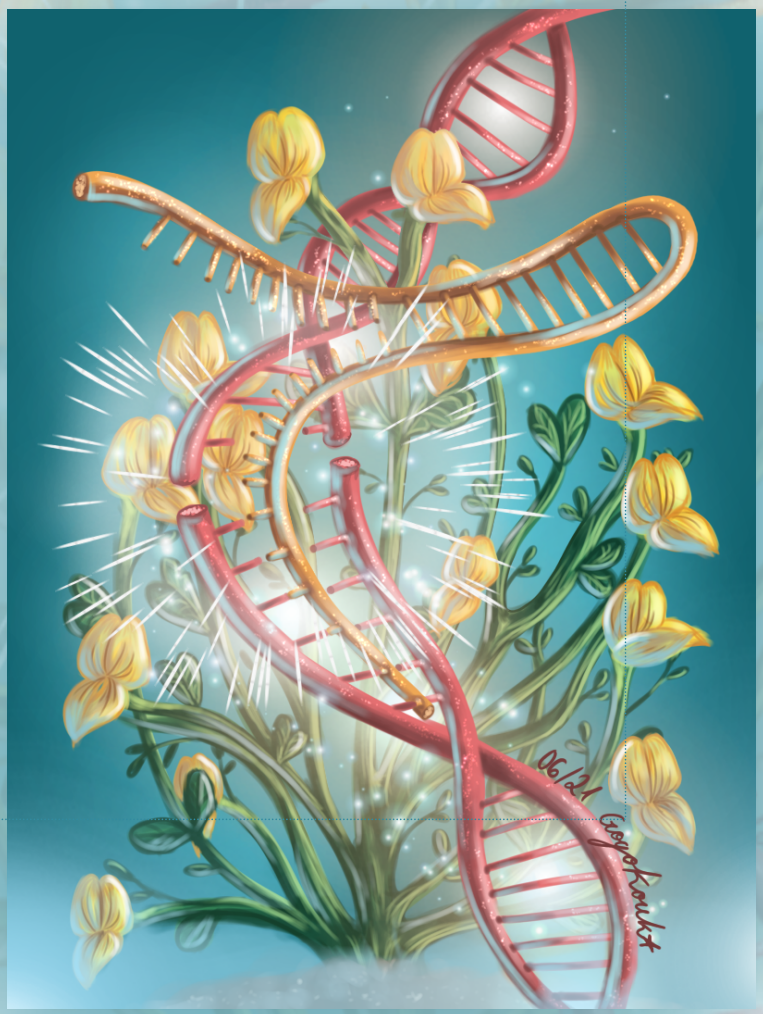
Department of Biochemistry and Biotechnology
University of Thessaly

Master of Science program:
"Advanced Experimental & Computational Biosciences"

# *Genome editing in L. japonicus using CRISPR/Cas9 technology*

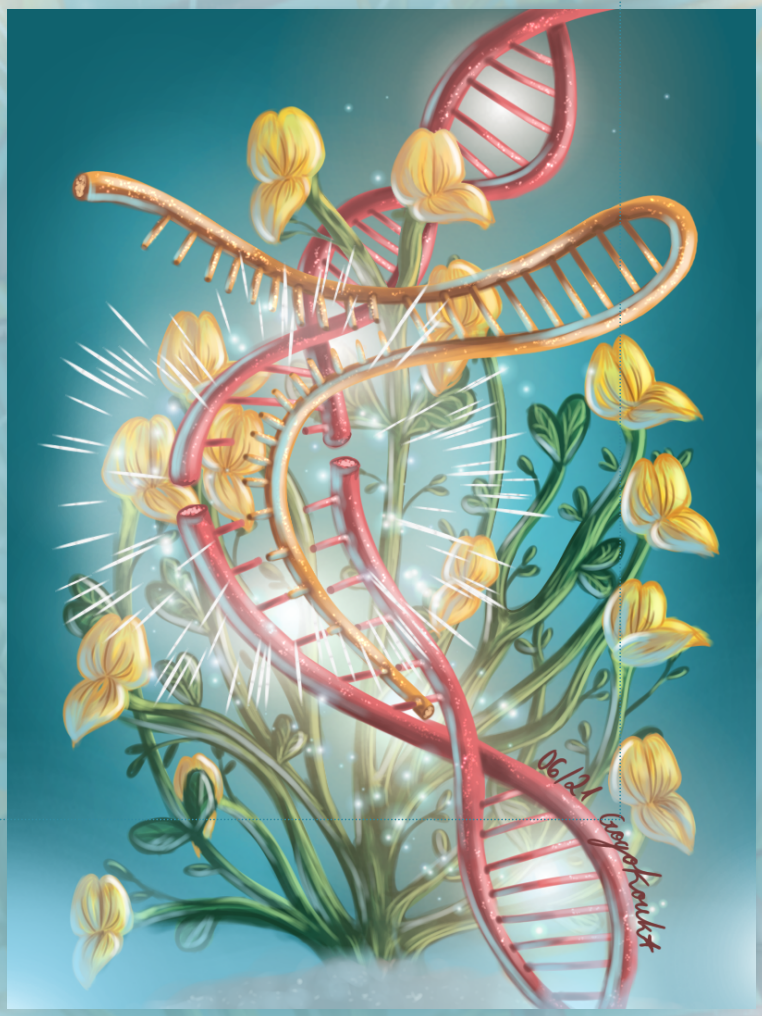



Master Thesis
Koukara Evgenia
June 2021

Τμήμα Βιοχημείας και Βιοτεχνολογίας
Πανεπιστήμιο Θεσσαλίας

Μεταπτυχιακό πρόγραμμα σπουδών
"Προηγμένες Πειραματικές και Υπολογιστικές Βιοεπιστήμες"

# *Γονιδιωματική τροποποίηση του φυτού L. japonicus μέσω της τεχνολογίας CRISPR/Cas9*




Μεταπτυχιακή Διατριβή
Κουκάρα Ευγενία
Ιούνιος 2021

## Three-member committee

**Papadopoulou Kalliope**

Professor of Plant Biotechnology

**Giakountis Antonios**

Assistant Professor of Molecular Biology-Genomics

**Tsikou Daniela**

Assistant Professor of Molecular and Developmental Plant Biology

# Aknowledgements

*I am deeply grateful to my supervisor Pr. Kalliope Papadopoulou for all of her help and support over the few last years, within and outside of the lab. I wish to thank the present and past members of the research group of Plant & Environmental Biotechnology for their help and support. Particularly, i would like to aknowledge Dr Constantine Garagounis, PhD candidate Magdaline Tsitsikle and MSc student Nikolaos Delkis for our collaboration through the conduction of my bachelor and master thesis.*

*I would like to express my gratitude to Pr. Kostas Mathiopoulos for co-creating this Master of Science program. We learned and grew a lot, professionally and mentally, through out the two-year duration. Special thanks to all members of the program, notably to Pr. Antonis Giakountisa and Pr. Dianiela Tsikou for advising me during the writing of this master thesis and during the lessons of the program.*

*Finally, I would like to thank my niece, Georgia Koukara, for the amazing painting of "Genome ediing in L. japonicus" depicted in the cover page.*

# Contents

Abstract……...6
Introduction……... 7
1. The model legume *Lotus japonicus*……7
2. Nodulation ……8
2.1 Legume-rhizobia perception and early signaling……...8
2.2 Rhizobial infection and nodule organogenesis…...……...8
3. Plant triterpenoid metabolism……...9
3.1. Triterpenoid metabolism in legume *L. japonicus*……..11
4. Lotus SHAGGY-like kinase 1……13
5. CRISPR/Cas9 technology; from an adaptive immune system to genome editing tool……...14
5.1. Discovery of CRISPR/Cas system……....14
5.2. CRISPR/Cas immunity……..14
5.3. CRISPR/Cas9-mediated targeted genome editing……15
5.4. CRISPR/Cas9-mediated genome editing in plants……17
Aim of this Master thesis…… ……21
Materials and methods……..22
1. Heat shock transformation protocol with DH5a chemo-competent bacterial cells……22
2. Golden Gate-based plasmid construction……22
2.1. Level 0 modules……23
2.2. N' SV40 NLS ds oligo……...26
2.3. Level 1 transcriptional units……...26
2.4. gRNA cassettes……....28
2.4.1. Selection of target sites and design of gRNA sequences……....28
2.4.2. Annealing of oligonucleotides……29
2.4.3. Cloning of dsoligos into gRNA acceptors……..30
2.5. Level 2 vectors……30
3. *Agrobacterium rhizogenes*-mediated hairy root transformation……....32
3.1. *L. japonicus* seed scarification/sterilization and germination……....32
3.2. *Agrobacterium rhizogenes* transformation and culture preparation....……32
3.3. Hairy root transformation……..…33
3.4. Identification and sampling of GFP positive transgenic hairy roots……...33
3.5. Genomic extraction from root samples using CTAB protocol……....33
3.6. Mutation analysis……33
3.6.1. PCR assay……...33
3.6.2. T7 endonuclease I assay……34
Results……..36
1. Assembly of a Golden Gate cloning toolkit optimisedfor gene editing in *L. japonicus*……36
1.1. CRISPR/Cas9 level 0 collection of stardard parts……36
1.2. CRISPR/Cas9 level 1 collection of transcriptional units……..43
2. CRISPR/Cas9 knockout of *L. japonicus* AMY2, CYP71D353 and LSK1 genes……..48
2.1. Selection of target sites ……48
2.2. Construction of gRNA expression cassettes……49
2.3. Assembly of complete CRISPR/Cas9 binary vectors……...50
2.4. *Agrobacterium rhizogenes*-mediated hairy root transformation……...52
2.5. Detection of target mutations……54
Discussion-Upcoming experimens……...59
Bibliography……61

# Abstract

The model legume *L. japonicus* has been explored for its ability to participate in beneficial symbiotic relationships and synthesise a plethora of triterpenes, a category of specialised metabolites. Closer investigation has revealed two interlaced associations; the first between triterpene metabolism, nodulation and a family of *GSK3b/SHAGGY*-like kinases, and the second between tritererpene metabolism and *L. japonicus*-FsK symbiosis. The *AMY2* gene cluster and *CYP88D4*, which create a new biosynthetic pathway by synthesizing b-amyrin and three other novel triterpene structures, are involved in regulating different aspects of nodulation and FsK colonization. *LSK1*, a *L. japonicus Shaggy*-like kinase, modulates nodulation and potentially interacts with lupeol, another primary triterpene. In order to elucidate these interconnections, we need a series of mutants depleted of specific triterpenes or LSK1 function. We decided to employ a CRISPR/Cas9 system optimisedfor mutagenesis in *L. japonicus* and generate a series of *amy2*, *cyp71d353* and *lsk1* stable silenced plant lines. We adopted a Golden Gate-based cloning strategy for our CRISPR/Cas9-based genome editing system and started by creating a CRISPR/Cas9 level 0 collection of standard parts. At the second phase, a series of CRISPR/Cas9 level 1 transcriptional units was generated based on the order and orientation of the expression cassettes in the final CRISPR/Cas9 level 2 vector. We designed multiple gRNAs and assembled several CRISPR/Cas9 binary vectors targeting *AMY2*, *CYP71D353* and *LSK1* separately. Subsequently, we performed a series of *A. rhizogenes*-mediated hairy root transformations to assess the mutation efficacy of each gRNA. In the future, after confirming which gRNAs are functional, we will proceed with the *A. tumefaciens*-mediated stable transformation to acquire *amy2*, *cyp71d353* and *lsk1* mutant lines. Our aim is to utilize the CRISPR/Cas9-mediated mutant seeds for functional analysis and understand better the interconnection between triterpene metabolism, endophytic interactions and *LSK1*.

# Introduction

## 1. The model legume *Lotus japonicus*

Legumes (Fabaceae or Leguminosae) constitute one of the largest families of flowering plants, whose members exhibit great agronomic and industrial significance. They are accountable for a large share of the world's primary crop production, being a highly nutritious source of plant-based proteins and essential amino acids. They are also used as fodder for animals (Graham & Vance, 2003). Furthermore, because of their secondary metabolism, they produce specialised metabolites (e.g. terpenes) with favourable properties for exploitation in industry and medicine (Laszczyk, 2009; Moses, Pollier, Thevelein, & Goossens, 2013). Another key aspect of legume biology is their ability to participate in beneficial symbiotic relationships with endophytes, the genus *Rhizobium* being one of them. This interaction accommodates the symbiotic nitrogen fixation, which provides these plants and subsequent crops with a free source of nitrogen (Smil, 1999).

Legume research aims to elucidate global and legume-specific physiological phenomena and biochemical pathways. On the other hand, many breeding programs strive to understand how different biotic and abiotic factors affect legume productivity and how to improve it (Graham & Vance, 2003). The legume *Lotus japonicus* (*L. japonicus*) is one of the two primary model legumes used to investigate the genetic background of legume-specific phenomena. *L. japonicus* (Regel) K. Larsen was first recorded at the ancient capital Kyoto of Japan. Its natural habitat is located in East- and Central Asia. Two cultivars are extensively used for research, *L. japonicus* Gifu B-129 and Miyakojima (MG-20), named after the place they were originally discovered. *L. japonicus* has been chosen as model species, because it exhibits favourable characteristics towards genetic, biochemical and functional manipulation. It is an autogamous diploid legume with a small genome size (~470Mb), organized in six chromosomes (2n=12). It has a short life cycle (2-3 months), producing a high number of seeds, thus facilitating the course of many experiments. Furthermore, it is amenable to plant transformation with *Agrobacterium rhizogenes* and *Agrobacterium tumefaciens*, while regeneration of whole plants is possible from calluses. Finally, *L. japonicus* synthesises a plethora of specialised metabolites, like triterpenes, while is capable of establishing rhizobial, mycorrhizal and other symbiotic relationships (Handberg, K. and Stougaard, J., 1992). One such symbiosis is with *Fusarium solani* strain K (**FsK**).

*Fusarium solani* is a species complex belonging in the Ascomycota phylum of the Fungi kingdom. Many members act as plant pathogens affecting important agricultural crops (Coleman, 2015). *Fusarium solani* strain K is a non-pathogenic endophyte, previously isolated from the roots of tomato plants (Kavroulakis et al., 2007). FsK is effective in protecting tomato plants against pathogens of the root and leaf system (Kavroulakis et al., 2007), spider mites (Pappas et al., 2018) or zoophytophagous insects (Garantonakis et al., 2018). Finally, FsK has been shown to increase plant tolerance to drought (Kavroulakis et al., 2018). Apart from tomato-FsK symbiosis, FsK is able to efficiently colonize the roots of the two model legumes *L. japonicus* (Fig. 1) and *M. truncatu*la without apparent symptoms. What is most certainly intriguing is the fact that FsK enters the plant through the roots, progressively colonizes the *L. japonicus* root system and subsequently migrates to the stem, in contrast to tomato plants, where FsK colonization is restricted to the root system (Kavroulakis et al., 2007; Skiada et al., 2019).

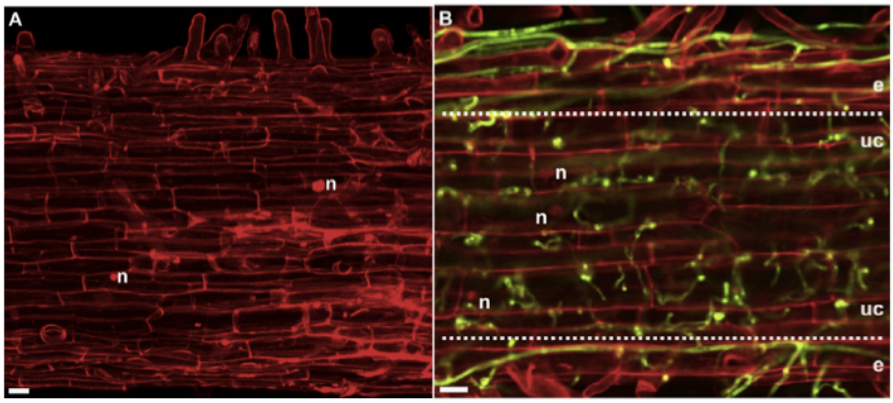


**Figure 1. Colonization of *L. japonicus* roots by FsK.** Confocal microscope images of FsK expressing cytoplasmic GFP and wild-type *L. japonicus* roots stained with Propidium Iodide. A. Control root tissues. B. FsK-expressing cytoplasmic GFP colonizes epidermal and cortical cells at 3 days post inoculation. (Adapted from Skiada et al., 2019).

Legumes are capable of participating in N-fixing symbiosis with soil rhizobia, when they grow in soils lacking mineral and organic N. Symbiotic nitrogen fixation not only supplies them with fixed N originating from atmospheric nitrogen ($N_2$), but also allows them to inject portions of the fixed N into the soil, hence providing natural and agricultural ecosystems with a free and harmless source of N. In return, symbiotic rhizobia gain a study supply of carbon (Boogerd & van Rossum, 1997; Guinel & Geil, 2002). Numerous genes coordinate legume-rhizobial symbiosis and different plant mechanisms have been recruited to regulate and ensure minimum energy waste. The processes of nodulation and symbiotic nitrogen fixation have been extensively analysed in *L. japonicus.*

## 2. Nodulation

### 2.1. Legume-rhizobium perception and early signaling

The first step of a legume-rhizobium symbiosis includes the reciprocal recognition between legume and rhizobium. Legumes excrete host-specific flavonoids and isoflavonoids in the N-poor rhizosphere, where they attract compatible rhizobia in a chemotactic manner (Peters, Frost, & Long, 1986; Redmond et al., 1986). These molecules/signals trigger the activation of *nod* genes and the synthesis and release of rhizobia-specific lipochitooligosaccharides or nodulation (**Nod**) factors from the rhizobia (Denarie and Cullimore, 1993). Nod factors are responsible for inducing nodule development in legume roots, by binding on Nod factor receptors in the plasmatic membrane of the epidermal root hair cells. This interaction is achieved between compatible plant hosts and host-specific rhizobia. A compatible partner of *L. japonicus* is the *Mesorhizobium loti* (*M. loti*) R7A**,** a strain commonly used in the laboratory (Handberg, K. and Stougaard, J., 1992).

In *L. japonicus,* Nod factor perception involves *LjNFR1* and *LjNFR5* plant symbiotic signaling genes, which encode for receptor-like kinases with extracellular Lysin motif (LysM) domains) (Madsen et al., 2003; Radutoiu et al., 2003). Symbiotic receptor-like kinase LjSYMRK, a leucine-rich repeat receptor-like kinase, is another plasma membrane-bound protein that interacts with LjNFR5 and is involved in Nod factor recognition (Antolín-Llovera, Ried, & Parniske, 2014). LjSYMRK is the first factor that participates in the **common symbiotic pathway (CSP)**. This pathway consists of a series of proteins and signaling molecules that act downstream of both fungal and rhizobial signal perception and activate the appropriate response for each symbiotic partner (Oldroyd, 2013). Three nucleoporins, NYP85, NYP133 (Kanamori et al., 2006; Saito et al., 2007) and NENA (Groth et al., 2010), localized at the nuclear pore complex, have been found to be essential for signal transduction downstream of LjSYMRK. They are postulated to be involved in transport or localization of the factor(s) needed for the induction of nuclear $Ca^{2+}$ spiking. Their precise role in the symbiotic signaling pathway remains to be elucidated (Kouchi et al., 2010). A central event of CSP that takes place within the nucleus is the repeated oscillations of $Ca^{2+}$ concentration, also known as spiking. LjCASTOR and LjPOLLUX (Imaizumi-Anraku et al., 2005; Charpentier et al., 2008), two cation channels located at the nuclear envelope, assist the sharp fluctuation of nuclear $Ca^{2+}$ concentration, while the main players remain to be identified. $Ca^{2+}$ spiking, in combination with calmodulin, is allegedly responsible for activating nuclear calcium/calmodulin-dependent protein kinase (LjCCaMK) through an intricate conformational change (Tirichine et al., 2006). The activated LjCCaMK phosphorylates transcription factor LjCYCLOPS, which in turn will regulate downstream nodulation specific regulators (Singh, Katzer, Lambert, Cerri, & Parniske, 2014).

Rhizobium perception activates two plant responses; rhizobial infection, more commonly through formation of infection threads at the epidermal root hairs and nodule organogenesis through formation of nodule primordia at the inner root cortex.

### 2.2. Rhizobial infection and nodule organogenesis

In *L. japonicus,* the most common and efficient mode of rhizobial entry into root cells is the formation of tubular structures, called infection threads (**ITs**), in the root hairs (Fig. 2) (Sprent & James, 2007). This entry

route presupposes the attachment of rhizobia on root hairs and activation of early nodulation-signaling pathways, which lead to root hair curling. As a result, rhizobia are entrapped within the newly created cavity, referred to as the infection pocket, where they will multiply and create micro colonies. The construction of the infection thread starts from the infection pocket, where the cell wall is deconstructed and the plasma membrane invaginates towards the root centre to allow the entry of rhizobia into the plant (Gage and Margolin 2000).

While the rhizobial infection process starts at the epidermal root hair cells, nodule organogenesis is concomitantly initiated in the root centre (Fig. 2). Differentiated cells at the root cortex and pericycle start to divide to develop a nodule meristem. Infection thread formation progresses until the nodule meristem, where rhizobia will be released into the cytoplasm of the nodule cells. Initially, bacteria inhabit the plant cells enclosed by the peribacteroid membrane, derived from the plant plasma membrane. Later, they will differentiate into the bacteroid, their symbiosis-specific form, followed by developing nitrogen-fixing activity. At their final maturation stage, propagation is ceased and the peribacteroid membrane-enclosed bacteroids, now referred as 'symbiosome', are actively fixing atmospheric nitrogen (Oldroyd, Murray, Poole, & Downie, 2011). It is considered that there are many host plant regulatory mechanisms regulating bacteroid differentiation and nitrogen fixation, such as communication between uninfected and infected nodule cells.

Nod factor perception by LjNFR1 and LjNFR5, as well as bacterial surface signal perception by other plant plasma membrane receptors, activates two spatially different physiological responses, which are, however, temporally regulated. The coordinated development between infection progression and nodule organogenesis is considered to be crucial for efficient nitrogen-fixing nodule formation (Oldroyd and Downie, 2008). The initial stages of legume-rhizobium symbiosis, as well as later stages, like nodule organogenis and symbiotic nitrogen fixation, are closely monitored by plant endogenous mechanisms, including hormonal regulation, autoregulation of nodulation (AON), **families of signaling proteins** and **triterpene metabolism**.

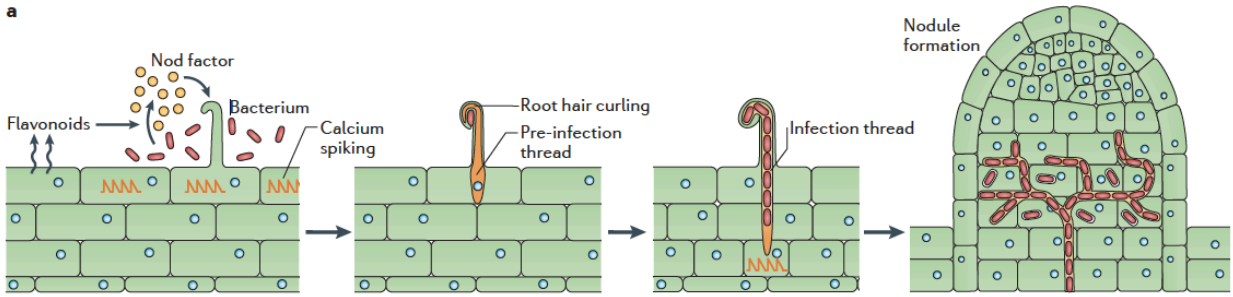



**Figure 2**. **Nodulation comprises two spatially different but temporally regulated plant responses.** Flavonoids released by the plant root attract compatible rhizobia in a chemotactic manner. In return, rhizobia release specific lipochitooligosaccharides or nodulation (**Nod**) factors. Attachment of rhizobia onto root hairs and nod factor perception activate early nodulation-signaling pathways (via calcium oscillations), which lead to root hair curling. As a result, rhizobia are entrapped within the newly created cavity, referred to as the infection pocket, where they will multiply and create micro-colonies. The construction of the infection thread starts from the infection pocket, where the cell wall is deconstructed and the plasma membrane invaginates towards the root centre to allow the entry of rhizobia into the plant cells. In the mean time, differentiated cells at the root cortex and pericycle start to divide to develop a nodule meristem. Infection thread formation progresses until the nodule meristem, where rhizobia will be released into the cytoplasm of the nodule cells. Rhizobia are released into membrane-bound compartments within nodule-specific cells and differentiate into a nitrogen-fixing state. Adapted from (Oldroyd, 2013).

## 3. Plant triterpenoid metabolism

Plants are notorious for their ability to synthesise primary and specialised (previously secondary) metabolites. While primary metabolites are ubiquitously expressed and characterized as essential for basic processes (e.g. respiration and photosynthesis), specialised metabolites were originally considered to perform only ecological functions. Being species- or taxa- specific, their role was restricted to increasing the survival rate of plants growing in adverse environments by enabling them to adapt more efficiently (Erb & Kliebenstein, 2020). However, recent studies ascribe novel functions to them. For instance, oats utilize the b-

amyrin-derived triterpene glycoside avenacin A-1 to defend themselves against fungal pathogens in the root system (Papadopoulou, Melton, Leggett, Daniels, & Osbourn, 1999). Kemen et al showed that accumulation of the basic triterpene b-amyrin in oat roots affected the root epidermal cell patterning (Kemen et al., 2014). These studies indicate that triterpenes may be involved in different aspects of plant physiology.

Triterpene biosynthesis begins with the selective cyclization of 2,3-oxidosqualene into various triterpene scaffolds by oxidosqualene cyclases (**OSCs**). B-amyrin synthase, a-amyrin synthase and lupeol synthase synthesise b-amyrin, a-amyrin and lupeol, respectively (Abe, Rohmer, & Prestwich, 1993; Abe, 2007), following the chair-chair-chair (CCC) conformation (Xu, Fazio, & Matsuda, 2004). The basic backbone of these triterpene scaffolds is further modified via site-specific oxidation by cytochrome P450 monooxygenases (**CYPs**) (Fig. 3). The palette is further enriched, since some of the produced triterpenoid aglycones are glycosylated into triterpenoid saponins by UDP-dependent glycosyltransferases (UGTs) and cellulose synthase-like enzymes (Chung et al., 2020).

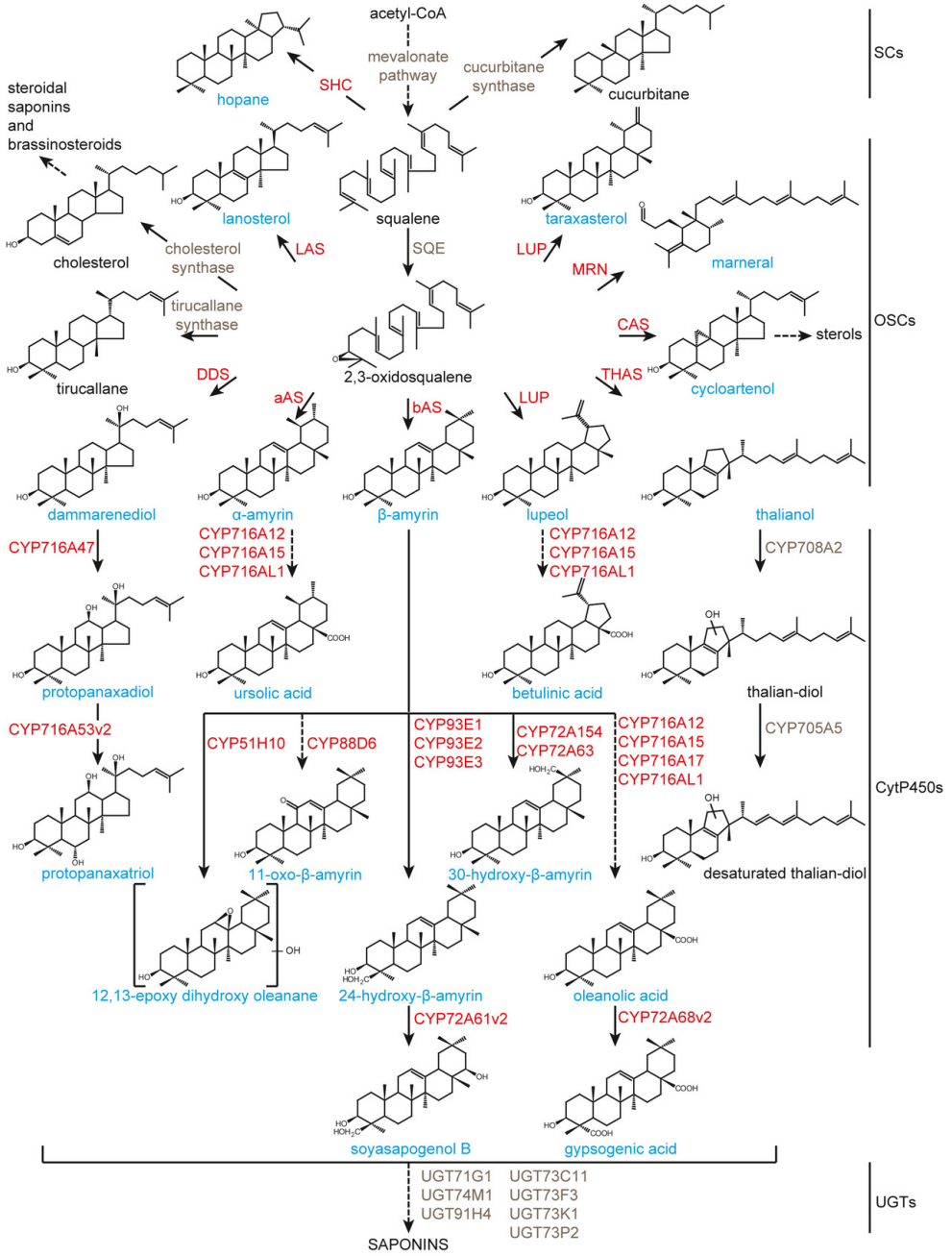


**Figure 3. Plant triterpenoid metabolism includes a variety of different structures synthesised by multiple gene families.** Dotted arrows indicate multiple steps. Enzymes are highlighted in red, compounds in blue. Adapted from (Moses, Pollier, Thevelein, & Goossens, 2013).

### 3.1. Triterpenoid metabolism in legume *L. japonicus*

Triterpene biosynthesis has also been investigated in *L. japonicus*. Four oxidosqualene cyclases (OSCs) and five cytochrome P450 monooxygenases (CYPs) have been characterized via molecular cloning of cDNA and heterologous expression in yeast or *in planta*.

***LjAMY2*** (OSC8) was the first triterpene biosynthetic gene to be partially characterised in *L. japonicus* (Iturbe-Ormaetxe, Haralampidis, Papadopoulou, & Osbourn, 2003). It is defined as a multifunctional b-amyrin synthase, because it synthesises two triterpenes, b-amyrin being one of them. The other triterpene scaffold is a novel structure, with an open ring, that is sequentially modified by ***LjCYP71D353***, ***LjCYP88D4*** and ***LjCYP88D5*** (Fig. 4). *AMY2* and these three *CYPs* create a new triterpene biosynthetic pathway, whose genes are tightly co-regulated during root and nodule development. Furthermore, all four genes are localised in the chromosome three, in close proximity to each other, while *AMY2, CYP71D353* and *CYP88D5* create a metabolic **gene cluster** (Fig. 5). Co-expression analysis revealed that the *AMY2* gene cluster and *CYP88D4* respond to *L. japonicus* inoculation with *M. loti* strain R7A (Krokida et al., 2013; unpublished results).

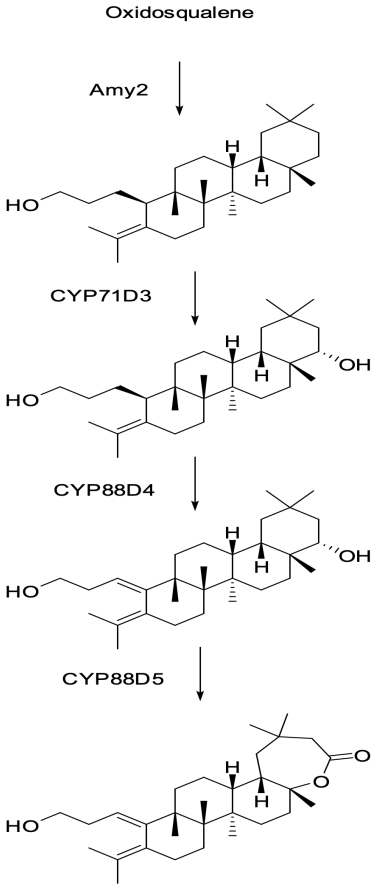



**Figure 4. Proposed novel triterpene pathway of *AMY2* gene cluster and *CYP88D4*.** Compound structures were determined via NMR (unpublished results).

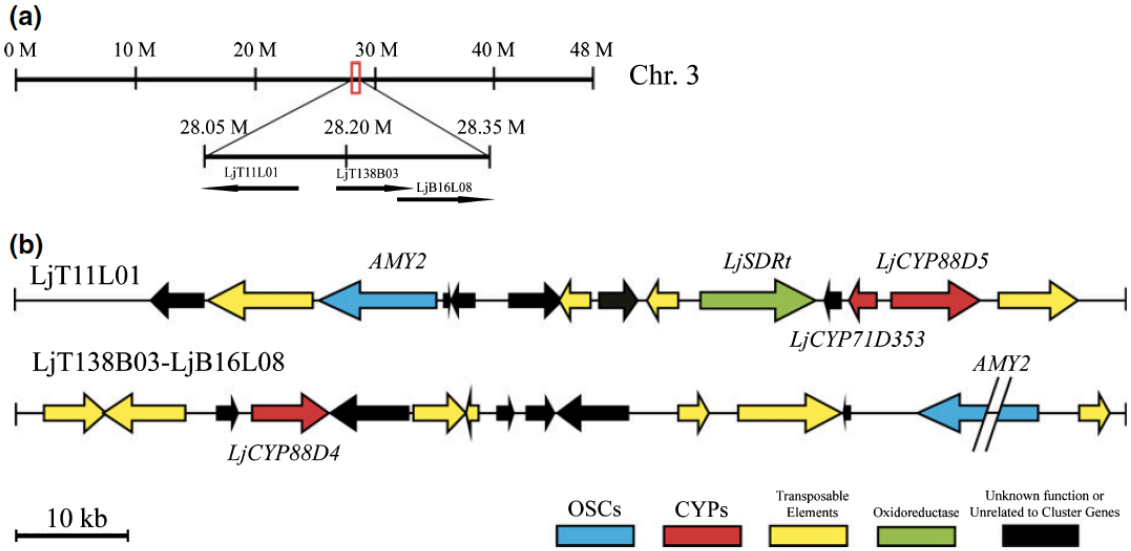



**Figure 5. *AMY2* gene cluster in *Lotus japonicus*.** Three genomic clones revealed the presence of two triterpene clusters. The first cluster consists of the *AMY2*, *CYP71D353*, *CYP88D5* and *SDRt* genes (LjT11L01). Expression and functional analysis revealed that *AMY2*, *CYP71D353* and *CYP88D5* are involved in triterpene biosynthesis. The second cluster comprises an *AMY2*-interrupted gene, which is most likely not functional, and *CYP88D4* (LjT138B03-LjB16L08).

To investigate the putative role of the AMY2 pathway in regulating the processes of host-rhizobium interaction, nodulation and symbiotic nitrogen fixation, a set of different gene silencing methodologies has been employed, with some inconsistencies between the observed phenotypes. No obvious effects were observed with regard to the nodulation process (i.e. nodule number at 20 and 40 days post inoculation) in RNAi-silenced *AMY2*, *CYP71D353* or *CYP88D5* trangenic hairy roots. Only a more severe retardation of the rate of hairy root growth was observed in 40 dpi than 20dpi compared to control trangenic hairy roots. In contrast, RNAi-*AMY2* stable silenced plants exhibited a short, stunted root phenotype and did not flower (Krokida et al., 2013). Because of the inconclusive results, another *AMY2* mutant was utilized, amy2-188. This was an EMS tilling mutant, harbouring a substitution from guanine to adenine that caused an early termination codon. *AMY2* silencing positively regulated infection progression, size and maturity rate of nodules. Interestingly enough, cyp71_N30 plants (*CYP71D353* silenced plants through transposon insertion) or RNAi-silenced *CYP88D4/D5* plants exhibited not only enhanced infection progression, but also decreased total nodule number (unpublished results). From the last results, it seems that certain intermediate or end products of the *AMY2* pathway regulate different aspects of nodulation. The above observations should be evaluated in other plant mutant lines to assess reproducibility of the observed phenotype. Additionally, more physiological characteristics of the mutant lines need to be evaluated, microscopically and macroscopically, to clarify which symbiotic nitrogen fixation- related processes are affected and in what stage.

Early findings implicate the AMY2 pathway in another *L. japonicus*-endophyte interaction with FsK. *AMY2* expression is induced in FsK-treated plants at early stages. This induction seems to negatively regulate FsK colonization, since *amy2* or even *cyp71d353* mutants exhibit increased colonization levels (unpublished results).

*OSC3* is another oxidosqualene cyclase encoding for a lupeol synthase (Sawai, 2006). *OSC3* is located on chromosome two. *OSC3* expression is induced in roots and nodules of *M. loti*-treated plants (Delis et al., 2011) and is coexpressed with *CYP716A51* (localized in chromosome four) in roots of uninoculated plants. Betulinic acid, the product of sequential enzymatic action of *OSC3* and *CYP716A51* (Fig. 6), is highly accumulated in the roots (Suzuki et al., 2019). Lupeol, or its derivative, seems to negatively regulate nodule organogenesis rate. RNAi-silenced *OSC3* roots exhibited increased number of nodules at 20 days post inoculation (dpi). This effect was restricted at early stages, since silenced and wild type plants showed similar numbers of nodules at 40dpi. Additionally, the expression of *ENOD40*, a marker gene for nodule primordia initiation, was increased in the *OSC3* silenced plants (Delis et al., 2011). *AMY1* (*OSC1*) is another oxidosqualene cyclase responsible

for synthesising b-amyrin in *L. japonicus* (Sawai, 2006). *AMY1* is located on chromosome three. In contrast to *AMY2*, *AMY1* is expressed in other tissues, other than roots, and is co-expressed with other genes, presumably composing a triterpenoid biosynthetic pathway. According to available RNA-seq data, their expression was affected in *M. loti* infected plants (unpublished results). However, the role of *AMY1* has not yet been extensively analyzed in *L. japonicus*. Lastly, *OSC9* is an a-amyrin synthease that is also co-expressed with *CYP716A51*, mainly in the leaves and stems (Suzuki et al., 2019). Its role in *L. japonicus* physiology remains to be characterised.

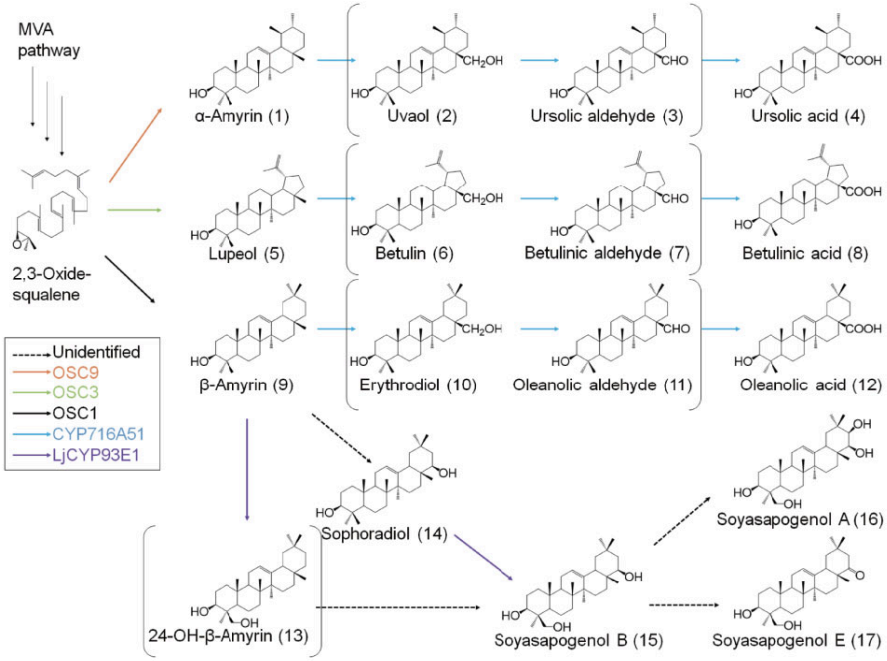



**Figure 6. Triterpene metabolism in *L. japonicus***. AMY1, OSC3 and OSC9 putative triterpene pathways are depicted. Adapted from (Suzuki et al., 2019).

Since AMY2, OSC3 and AMY1 pathways seem to be implicated in *L. japonicus*-endophyte interactions, we wished to generate a series of knock out mutants to investigate more closely the role of triterpene metabolism in said interactions. **We decided to first focus on the AMY2 gene cluster, since its silencing exhibits a series of interesting phenotypic characteristics in regards to nodulation.**

## 4. Lotus SHAGGY-like kinase 1

The family of GSK3b/SHAGGY-like kinases (SKs) consists of signaling proteins that participate in metabolic regulation, differentiation and other developmental processes in plants and animals (Claisse, Charrier, & Kreis, 2007; De Rybel et al., 2009). Plant SKs are also involved in hormone regulation and mediating biotic and abiotic stress responses (Jonak & Hirt, 2002; Youn & Kim, 2015). The role of SKs in legume physiology and particularly in regards to the nodulation has not been investigated, until recently.

Using a previously cloned SK cDNA as query, six sequences (*Lotus japonicus* SK1-6 or LSK1-6) were identified that contained Ser/Thr protein kinase domains, with conserved ATP-binding and kinase active sites. Gene expression analysis revealed that **LSK1** transcripts increased significantly in roots inoculated with *M. loti* R7A. Its putative role in nodulation was further probed by functional analysis in *A. rhizogenes*-mediated hairy root transformation and mutant plants. RNAi-silenced *LSK1* roots and a homozygous transposon-tagged insertion line lsk1-1 exhibited increased number of nodules compared to wild type inoculated plants. LSK1-overexpressing hairy-root lines did not exhibit any statistically significant alteration to the nodule number (Garagounis et al., 2019).

Since the functional characteristics were unaffected in *LSK1* silencing and overexpressing lines, the

activation of the autoregulation of nodulation (AON) pathway was assessed. This mechanism is activated after rhizobium infection and suppresses nodulation in response to sufficient nitrate levels. The expression of AON genes was affected in *LSK1* knockdown plants at low nitrate levels, both at early and late stages of nodulation. At higher levels of nitrate, the expression pattern of AON-related genes was reversed compared to that at low levels of nitrate, at early stage of nodulation. Interestingly enough, *LSK1* transcript levels did not change in response to high levels of nitrate in either wild type or *lsk1-1* plants at early stages of nodulation (Garagounis et al., 2019). In conclusion, *LSK1* modulates the expression of AON components in response to rhizobium inoculation differently at low and high levels of nitrate independent of its transcription expression levels.

*Lotus japonicus* employs an additional regulatory mechanism that interconnects nitrogen levels with the AON pathway in order to achieve balance between nodule formation and energy waste. This intertwined network of regulatory mechanisms is possibly further enriched by the involvement of the triterpenoid metabolism. In vitro experiments have shown that lupeol binds to *LSK1* (unpublished results). In order to study the relationship between *LSK1*, triterpenoid metabolism and nodulation, we need a variety of stable silenced plant lines that are easily generated and utilised, without the risk of introducing unwanted mutations during each generation. **We decided to employ the CRISPR/Cas9 technology to create *amy2*, *cyp71d353* and *lsk1* stable silenced plant lines and analyse their loss-of-function phenotypes in future experiments.**

## 5. CRISPR/Cas9 technology; from an adaptive immune system to genome editing tool

### 5.1. Discovery of CRISPR/Cas system

Initial findings for the existence of CRISPR/Cas systems were discovered by chance in *Escherichia coli* strain K12. Ichino and colleagues, while analyzing a chromosomal segment containing the alkaline phosphatase (iap) gene, identified a direct-repeated sequence of 29 nucleotides interspaced by a non-repeated sequence of 32 nucleotides (Ishino, Shinagawa, Makino, Amemura, & Nakata, 1987). A variety of interspaced short sequence repeats was also discovered in other bacterial and archaeal species. Later, it was shown that these motifs are members of the same family as they are composed of direct repeats, varying in size from 21 to 37 bp, interspaced by similarly sized non-repetitive sequences called spacers. Because of the common characteristics, this family was designated as **CRISPR**; clustered regularly interspaced short palindromic repeats. In silico analysis of the surrounding chromosomal region revealed the presence of the CRISPR-associated (***cas***) genes. This cluster is only present in CRISPR-containing prokaryotes and always adjacent to the CRISPR locus, suggesting a putative functional interaction. Both CRISPR and Cas loci are absent from eukaryotic and viral genomes (Jansen, Embden, Gaastra, & Schouls, 2002). Additional studies concluded the intervening, non-repetitive sequences to be of viral or plasmid origin, while a correlation between CRISPR locus and immunity was proposed (Bolotin et al., 2005; Mojica et al., 2005; Pourcel et al., 2005). In 2007, it was experimentaly shown that CRISPR locus, in combination with Cas genes, provided acquired resistance against phages in *Streptococcus thermophilus* (Barrangou et al., 2007). One year later, experiments revealed that CRISPR/Cas is also responsible for preventing conjugation and plasmid transformation in *Staphylococcus epidermidis* (Marraffini et al., 2008). Therefore, the biological role of CRISPR/Cas system is to initiate an adaptive immune response in prokaryotes (Barrangou et al., 2007). The involvement of different research groups was essential for elucidating the stages of immunity and characterizing the variety of natural CRISPR/Cas systems.

### 5.2. CRISPR/Cas immunity

Each CRISPR/Cas immunity is encoded by the complimentary actions of a CRISPR locus and a cluster of *Cas* genes. The CRISPR locus comprises a series of short DNA repeats separated by spacers of phage or plasmid origin and is preceded by a leader sequence encompassing its promoter. The cluster of *Cas* genes borders the CRISPR array and encodes proteins required for the immunisation and immunity stages of the CRISPR/Cas immunity (Fig. 7A). During the immunisation or adaptation phase, sequences from viral genome or plasmid are integrated into the CRISPR array (Barrangou et al., 2007). The immunity is subsequently expressed by the transcription of the CRISPR array into a long precursor CRISPR RNA (pre-crRNA) and further processing into small interfering CRISPR RNAs (crRNAs), each one containing one spacer sequence (Brouns et al., 2008). During the final stage of the immunity, mature crRNAs are used as guides to direct the cleavage of foreign homologous sequences, called protospacers, by Cas endonucleases (Garneau et al., 2010).

All CRISPR/Cas systems harbour the *Cas1* and *Cas2* genes that are central to the immunisation stage of the pathway, since they are involved in the acquisition of new spacers in the CRISPR array (Marraffini & Sontheimer, 2010). However, based on the accessory cas gene content, the mechanism of crRNA biogenesis and the targeting requirements, each system can be classified into three different types (I, II and III; Makarova et al., 2011). Type II systems have been extensively analyzed (Fig. 7B). These systems comprise a third genomic locus encoding for a small RNA known as trans-encoded crRNA (tracrRNA). TracrRNA is bound by **Cas9** endonuclease, the principal perpetuator of type II systems, and is partially complimentary to the repeat sequences of pre-crRNAs. CrRNA maturation requires cleavage of the repeat-tracrRNA dsRNA by the endogenous RNAse III (Deltcheva et al., 2011). After further processing, the mature crRNA-tracrRNA complex leads Cas9 endonuclease to complimentary regions of the foreign genome. Target recognition and melting of the two DNA strands requires the presence of a sequence downstream of the protospacer, referred to as protospacer adjacent motif (PAM). Cas9 endonuclease can recognise the 3-nt 5'NGG3', with N any nucleotide (Jinek et al., 2012). It can, however, also recognise 5'NAG3' (Hsu et al. 2013). The two catalytic domains of Cas9, RuvC and HNH, each of which cleaves one DNA strand of the protospacer region, accomplish target cleavage (Sapranauskas et al., 2011).

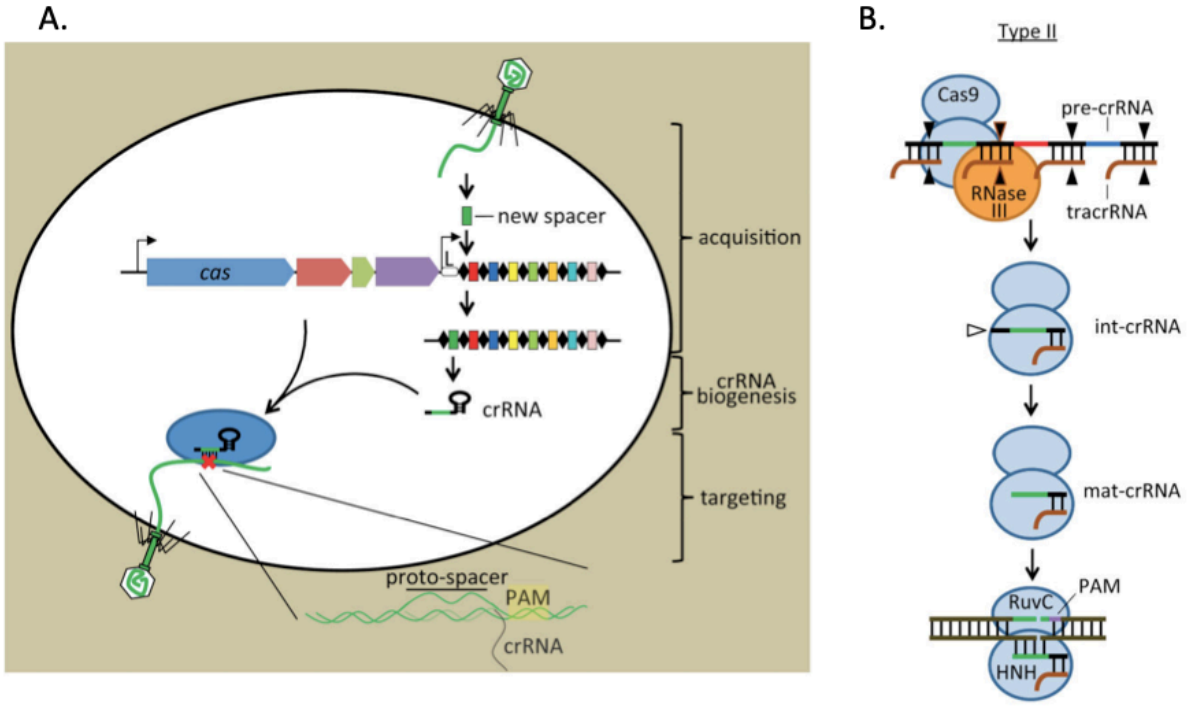



**Figure 7**. **Stages of CRISPR/Cas immunity**. **A)** A CRISPR locus contains an array of repeats (black diamonds) and spacers (coloured boxes) that are flanked by a "leader" sequence (L) and a CRISPR-associated (cas) gene cluster. During adaptation, new spacers derived from the genome of the invading virus are incorporated into the CRISPR array. The synthesis of a new repeat is also required. During crRNA biogenesis, a CRISPR precursor transcript is processed to generate small crRNAs. During targeting, mature crRNAs are used as guides to direct the cleavage of foreign homologous sequences, called protospacers, by Cas endonucleases. The existence of PAM site is prerequisite for cleavage of the foreign genome, while it is absent from the bacterial genome. **B)** In Type II systems primary processing of crRNAs requires the annealing of the tracrRNA to the repeat sequences of the pre-crRNA and the subsequent cleavage of the dsRNA by the host RNase III. After further processing, the mature crRNA-tracrRNA complex leads Cas9 endonuclease to complimentary regions of the foreign genome. Target cleavage requires the crRNA, the tracrRNA and the RuvC and HNH domains of Cas9, each of which leaves one DNA strand of the proto-spacer region, 3-nt upstream of the PAM site. Adapted from (Barrangou & Marraffini, 2014).

## 5.3. CRISPR/Cas9-mediated targeted genome editing

Targeted genome editing tools have been recruited in place of old traditional approaches such as random chemical mutagenesis. The principle of performing targeted genome editing is the generation of a DNA double-stranded break (DSB) at the genomic locus to be modified. The subsequent introduction of a mutation is possible due to the existence of two innate cellular pathways (Fig. 8) (Symington & Gautier, 2011). Non-homologous end-joining (NHEJ) pathway repairs DSBs by introducing nucleotide insertions or deletions (INDELS) of random length, which can disrupt the translational reading frame of a coding sequence or the binding sites of *trans*-acting factors in promoter or enhancer regions. NHEJ is by far the most common DSB repair mechanism in most organisms, including higher plants (Puchta, Dujon, & Hohn, 1996) (Puchta, 2004). Another repair mechanism is the homology-directed repair (HDR) pathway, which can be used to introduce specific point mutations or inserts of desired sequence through recombination of the target locus with the flanking arms of a "donor" template (Li et al., 2013).

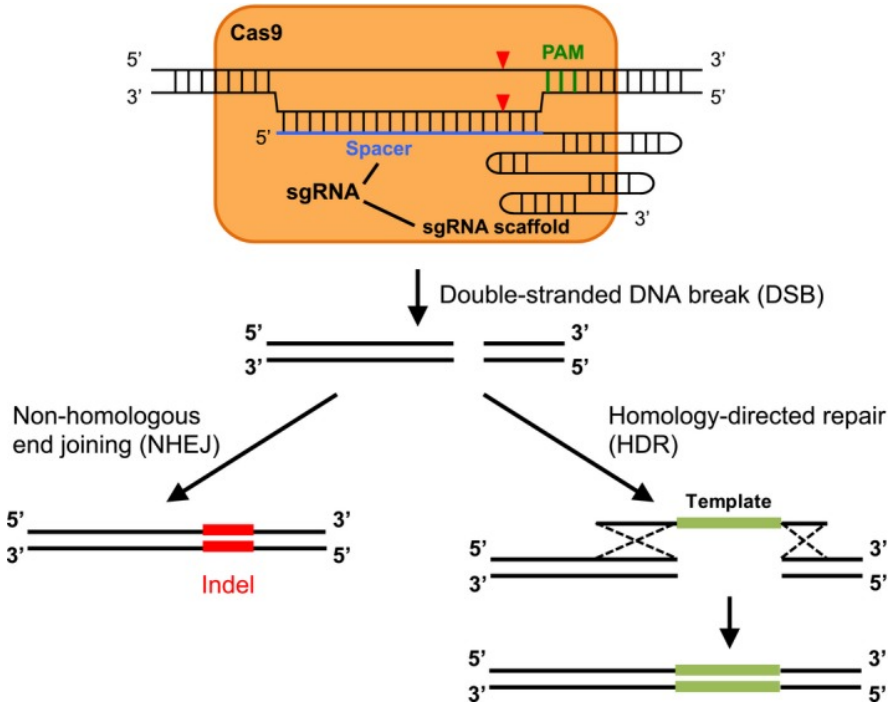



**Figure 8**. **The mechanism of CRISPR/Cas9-mediated genome editing**. The Cas9-sgRNA complex will bind to any genomic region with a PAM sequence (shown in green). If the spacer sequence (shown in blue) of the sgRNA is complementary to the strand harboring the complimentary sequence of the PAM site (3'NCC5'), Cas9 endonuclease will create a DSB three nucleotides upstream of the PAM site (indicated by red triangles). DSBs can be repaired by two pathways. The NHEJ pathway introduces indels at the DSB site. Alternatively, if a donor template is present, the HDR pathway can introduce specific point mutations or inserts of desired sequence through recombination of the target locus with the flanking arms of a "donor" template. Adapted from (Shan, Soltis, Soltis, & Yang, 2020).

Zinc fingers nucleases (ZFNs; Bibikova et al., 2002) and transcription activator-like effector nucleases (TALENs; Li et al., 2011) were the first nucleases to be used as genome editing tools. They are artificial chimeric proteins consisting of a DNA binding subunit fused to the non-specific nuclease subunit of FokI endonuclease. However, another more versatile nuclease-based genome editing tool came to replace them. Since natural CRISPR/Cas systems are based on the RNA-guided mediated cleavage of target sites, they were considered ideal targets for transitioning from natural defence systems to genome editing tools. Among various CRISPR systems, engineered class II CRISPR/Cas9 is the most popular and robust, especially Cas9 from *Streptococcus pyogenes* (SpCas9). It was shown that by simply changing the crRNA sequence, the target sequence could be reprogrammed. In addition, the 3' end of the crRNA could be fused to the 5' end of the tracrRNA creating a single chimeric RNA molecule, called single guide RNA or sgRNA or gRNA (Jinke et al., 2012). Hence, the system was simplified from three to two components, a Cas9 endonuclease and a sgRNA

molecule. The ribonucleoprotein Cas9-sgRNA complex binds any genomic region with a PAM sequence. If the spacer sequence (first 20 nucleotides at the 5′ end of the gRNA) is complementary to the genomic sequence immediately upstream of PAM, Cas9 endonuclease will make a DSB three nucleotides upstream of the PAM site (Fig. 8). The genome editing effect was first demonstrated in animals, like human cell lines, mouse and zebrafish. It was also possible to utilize multiple gRNAs with different sequences to achive multiplex genome editing at different loci simultaneously (Cong et al., 2013; Mali et al., 2013). Off-target effects consist a limiting factor that could deter CRISPR/Cas9 utilization. An off-target site is a sequence somewhere in the genome that shares high homology with that of the target CRISPR site (on-target site). If the PAM site is present, Cas9 endonuclease can cleave one or more off-target sites, apart from the on-target site, even if there are mismatches between the spacer and protospacer sequence. The position of mismatches can affect the off-target potential (Pattanayak et al., 2013; Fu et al., 2013). There are different CRISPR design tools that allow predict and minimize off-target effects.

### 5.4. CRISPR/Cas9-mediated genome editing in plants

The true potential of CRISPR/Cas9 technology for use in basic plant research and plant breeding programs was recognised and applied to a variety of plant species; model, ornamental and crop plants. The genome of model species *Arabidopsis thaliana* and *Nicotiana benthamiana* was targeted first, as well as other's species like rice (Feng et al., 2013; Li et al., 2013; Nekrasov et al., 2013; Xie and Yang, 2013). More plants were subsequently modified, including *Glycine max* (Jacobs, LaFayette, Schmitz, & Parrott, 2015), *Solanum lycopersicum* (Ron et al., 2014), *Zea mays* (Liang, Zhang, Chen, & Gao, 2014), etc. CRISPR/Cas9 has been successfully developed for use in polyploid species, including octoploid strawberry (Martin-Pizarro et al., 2018; Wilson et al., 2019) and tetraploid peanut (Yuan et al., 2019). CRISPR/Cas9-mediated gene editing has been established in some legume crops, including model legumes *Medicago truncatula* (Curtin et al., 2018) and *Lotus japonicus* (Wang et al., 2016, 2019) and legume crops such as soybean (e.g. Li et al., 2019), cowpea (Che et al., 2021) and chickpea (Badhan et al., 2021). **CRISPR/Cas9-mediated genome editing in *Lotus japonicus* can be accomplished with the following workflow.**

Selection of CRISPR/Cas9 target sites.

- CRISPR/Cas9-mediated genome editing in *Lotus japonicus* starts by identifying putative PAM sites on the plus and minus strand of the target genome locus. There are a variety of CRISPR design tools to choose from, but few of them offer the option to select *Lotus japonicus* as the reference genome. This step should be definitely taken into consideration, since prediction of off-target sites is crucial. Each tool uses its own algorithm to score putative on-target sites and potential off-target sites, hence selection of target sites based on different sources is recommended.

- If gene knock out is the desired outcome, targeting of coding region is the preferred method. Selection of target sites within the first exons or exons encoding for functional domains is encouraged. However, this is not always achievable, since many target genes belong to gene families whose members share high nucleotide similarity. In this case, to avoid undesired off-target sites, someone could choose target sites even in later exons or multiple target sites within less conservative intronic regions.

- For each target gene, two or more gRNAs should be designed that target different regions of the gene. Co-expression of multiple gRNAs can be achieved by connecting multiple gRNA cassettes tandemly in one plasmid construct (Li et al., 2013).

- Regarding off-target sites, off-target effects of CRISPR/Cas9 have been reported in plants, albeit at a low frequency. Nevertheless, they are less problematic in plant genome editing because off-target mutants can be discarded and undesired mutations can be eliminated through backcrossing (Belhaj et al., 2015).

Assembly of CRISPR/Cas9 vector

- A plasmid vector can be assembled harbouring the Cas9 expression cassette and one or multiple sgRNA (gRNA) expression cassettes. The selected cloning strategy should allow for efficient assembly of each CRISPR/Cas9 vector. Apart from the CRISPR/Cas9 components, additional cassettes can be included to facilitate effective selection or recognition of transgenic/transformed tissues.

Expression of sgRNA(s)

- RNA pol III dependent promoters such as U3 or U6 usually drive the expression of sgRNAs, since sgRNAs resemble small naturally encoded RNAs (Kumar & Jain, 2015). Notably, expression levels of sgRNA are higher when driven by endogenous promoters rather than exogenous ones (Sun et al., 2015).

- The identity of the first nucleotide at the 5' end of the produced sgRNA is designed based on the RNA pol III promoter driving its expression. For U6 promoters, preservation of G or A at or close to the first position close to the U6 promoter will increase transcription efficiency (Gao, Harwig, Berkhout, & Herrera-Carrillo, 2017). Taking this into consideration, sgRNA starting naturally with G or A can be selected or an additional nucleotide can be added during synthesis (Nakayasu et al., 2018).

- The length of gRNA sequence that is complimentary to the target genome sequence upstream of the PAM site should be also considered. 20nt is usually preferred, with an optional window between 17-22nt.

- Since RNA polymerase III uses a T-stretch as termination signal (consecutive Ts), a T-strech of equal or more than six consecutive Ts is used after the gRNA scaffold sequence to signal transcriptional termination. Keeping that in mind, T-stretch of at least four should be avoided within gRNA sequences (Gao, Herrera-Carrillo, & Berkhout, 2018).

Cas9 expression cassette

- Cas9 expression is usually driven by RNA pol II-dependent promoters such as ubiquitin gene promoters or a strong constitutive promoter like Cauliflower mosaic virus (CaMV) 35s promoter (Kumar & Jain, 2015).

- The sequence of SpCas9 should be codon-optimised based on which trinucleotides are preferred by the plant's translation machinery. So far, both dicotyledon codon-optimised Cas9 and human codon-optimised Cas9 have been successfully used in legume plants with relatively similar efficiency (Gasparis et al., 2018; Zhang, Yang, Yang, Li, & Guo, 2016).

- One or two nuclear localisation signals should be combined with the coding sequence of Cas9 to ensure proper translocation of protein into the nucleus (Belhaj, Chaparro-Garcia, Kamoun, & Nekrasov, 2013).

Delivery/transformation method

- There are various combinations of plant delivery and transformation methods based on the form of CRISPR/Cas9 system (cassettes or ribonucleoprotein complexes), plant species, tissues and experimental goal. For example, *N. benthamiana* leaf infiltration or protoplast transformation can be applied for assessing mutation efficacy of designed gRNAs (Li et a., 2013). *L. japonicus* is amenable to transformation using both *Agrobacterium rhizogenes* and *Agrobacterium tumefaciens*. Because *A. tumefaciens*-mediated stable transformation takes between 8 to 12 months to obtain seeds of transgenic plants, a transformation protocol using *A. rhizogenes* can be used instead to induce transgenic hairy roots. Composite plants with a transgenic root system can thus be used for complementation of plant mutants and other transgene studies (Handberg, K. and Stougaard, J., 1992). In our case, this approach can be employed to assess mutation efficacy of designed gRNAs and assembled CRISPR/Cas9 vector. After verifying mutations, we can move forward with the stable transformation. Another possibility is to perform functional analysis in transgenic edited hairy roots, if the phenotype of CRISPR/Cas9 gene silencing is visible in the root system.

- Either *A. rhizogenes* and *A. tumefaciens*-mediated transformation involves the transfer of a T-DNA region from the bacterial plasmid to the plant genome. Specific sites, called left and right border, delimit a T-DNA region. Therefore, plasmid vectors harbouring such sites are referred to as binary vectors (Bahramnejad, Naji, Bose, & Jha, 2019).

- The bacterial genes responsible for the T-DNA transfer are known as *vir* (virulence) genes. In nature, the

T-DNA region harbours genes that promote plant cell proliferation and synthesis of nutrients that are excreted and consumed by Agrobacteria. In the lab, Agrobacteria already encode vir genes endogenously and only transformation with a binary vector is required.

- *A. rhizogenes*-mediated transformation has some advantages over that with *A. tumefaciens*. Firstly, apart from the transfer of T-DNA harbouring the transgenes, another set of genes is transferred, referred to as rol genes. Plant cells transformed with the *rol* genes spontaneously differentiate into phenotypically transformed roots (hairy roots); an easily distinguishing feature. Secondly, hairy roots are genetically stable, as the transformed root presumably develops from a single transformed cell (Roychowdhury, Halder, & Jha, 2017).

Detection of mutations

There are different methodologies that can be used to assess mutation frequency (well-presented in the review by Shan, Soltis, Soltis, & Yang, 2020). The easiest would be to target a marker gene and assess the phenotype of the transformed tissues. Since this is not always the case or in order to confirm the gene editing event, the next step is to PCR amplify the targeted region including one or multiple target sites. Gel agarose electrophoresis can reveal indels of large size or even deleted regions between two target sites (Hashimoto, Ueta, Abe, Osakabe, & Osakabe, 2018). To search for indels of smaller size, PCR amplification can be combined with restriction enzyme assay or T7 endonuclease I assay.

PCR/RE assays take advantage of the presence of a restriction enzyme site within the CRISPR/Cas9 target site. After PCR amplification, amplicons from wild type alleles will be cleaved in the presence of the restriction enzyme. However, amplicons from modified alleles cannot be digested, since the recognition site has been destroyed by the NHEJ pathway. Therefore, gel agarose electrophoresis can reveal either three bands, with two corresponding to the digested wild type allele and one to the non-digested modified allele, or one band, if both alleles were modified (Cai et al., 2015).

PCR/T7 I assays recognise indels (but not SNPs). This approach is used when there is not any or common restriction enzyme sites close to the Cas9 cleavage site within the CRISPR/Cas9 target site. The targeted region is PCR amplified from the genomic DNA of the cell population, generating a mixture of amplicons. Some of them will originate from the wild type allele, while others from potential modified alleles. When the wild type and the modified PCR products denature and reanneal, mismatches and regions of single-stranded DNA are formed (heteroduplexes). T7 endonuclease I detects and cleaves most types of the single-stranded DNA regions. When run out on a gel, the full length of the wild type allele and cut products can be visualised and used as an approximation of editing efficiency (Cai et al., 2015).

To verify each mutation event, undigested PCR products can be sent directly for Sanger sequencing. A sequencing chromatograph with overlapping peaks close to the Cas9 cleavage site within the CRISPR/Cas9 target site suggests the existence of a mutation. The modified allele(s) can be differentiated from the wild type one using the Degenerate Sequence Decoding method (Ma et al., 2015). Alternatively, undigested PCR products can first be subcloned and multiple bacterial colonies can be Sanger sequenced to identify each mutation event. Finally, instead of Sanger sequencing, PCR products could be sent for NGS using an Illumina platform.

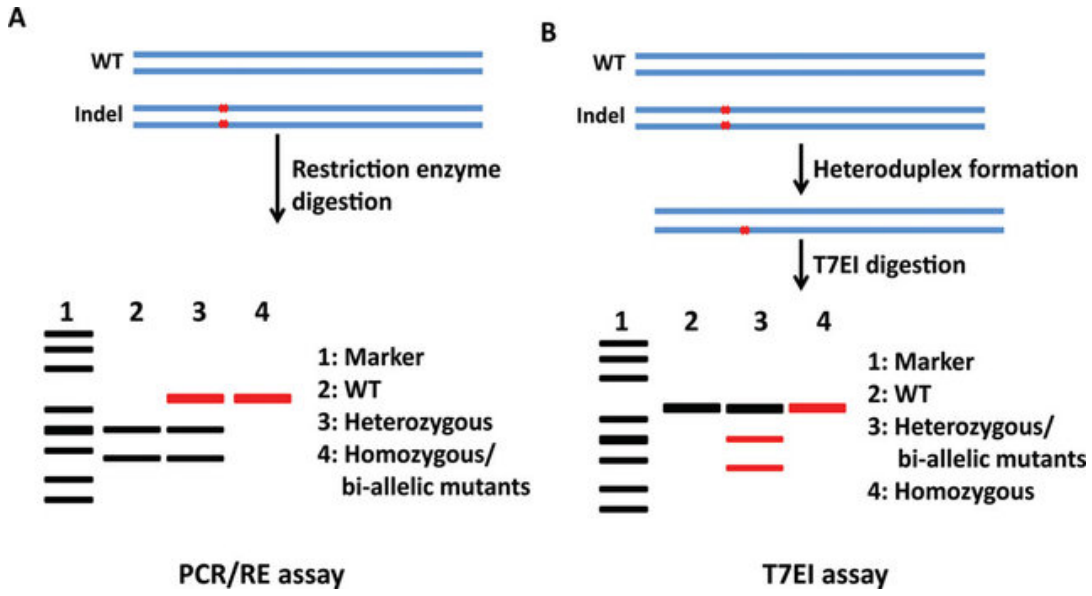



**Figure 9. Strategies to detect CRISPR/Cas9-mediated mutations**. (A) Schematic illustration of the PCR/RE assay. Wild type allele carries a restrictive digestion recognition site and amplicon is digested producing two fragments of smaller length. Heterozygous mutations can be partially digested and homozygous mutations and bi-allelic mutations can be completely undigested. (B) Schematic illustration of the T7EI assay. Heteroduplexes are produced if wild type and modified alleles are present in the genomic DNA. These are cleaved by T7EI. Homozygous mutants remain intact and cannot be distinguished from the wild type, since they cannot produce heteroduplexes. Adapted from (Liang, Zong, & Gao, 2016)

# Aim of this master thesis

Close investigation of the beneficial symbiotic interactions between model legume *L.* japonicus and rhizobium or FsK revealed a series of plant mechanisms at place to regulate them. On one hand, the AMY2 triterpene biosynthetic pathway is involved in regulating different aspects of nodulation and FsK colonization. On the other hand, *L. japonicus* Shaggy-like kinase 1 (*LSK1*) modulates nodulation and potentially interacts with lupeol, the product of triterpene biosynthetic *OSC3*. As we wish to understand how the *AMY2* pathway and *LSK1* affect these two symbioses, we decided to generate a series of CRISPR/Cas9-mediated *amy2*, *cyp71d353* and *lsk1* stable silenced plant lines. Because we adopted a Golden Gate-based cloning strategy for our CRISPR/Cas9-based genome editing system, we started by creating a CRISPR/Cas9 level 0 collection of standard parts followed by a series of CRISPR/Cas9 level 1 transcriptional units. Next, we designed multiple gRNAs and assembled several CRISPR/Cas9 binary vectors targeting *AMY2*, *CYP71D353* and *LSK1* separately. Subsequently, we proceeded with assessing the mutation efficacy of each gRNA through *A. rhizogenes*-mediated hairy root transformation. In the future, after confirming which gRNAs are functional, we will proceed with the *A. tumefaciens*-mediated stable transformation to generate *amy2*, *cyp71d353* and *lsk1* mutant lines and utilise them in functional analyses. Our ultimate goal is to extrapolate the acquired knowledge from *L. japonicus* to economically and environmentally important legumes.

# Materials and Methods

## 1. Heat shock transformation protocol with DH5a chemo-competent bacterial cells

Level -1, level 0, level 1 and level 2 vectors were propagated in *E. coli* DH5a chemo-competent cells. Two protocols were utilized, with the second optimisedfor efficient propagation of level 2 multigene vectors. Standard protocol started with addition of 5μl of digestion-ligation reaction to 80-100μl *E. coli* DH5a chemo-competent cells. After gentle mixing, cells were incubated with the cloning reaction on ice for 20 minutes. Heat shock was conducted at 42$^{o}$C for 55 seconds, followed by a 5-min incubation on ice. For the recovery stage, transformed cells were diluted in 1ml of pre-warmed LB liquid medium (10g tryptone, 10g NaCL and 5g yeast extract in diluted in 1L $dH_2O$, autoclaved at 121$^{o}$C, for 20 min) and incubated at 37$^{o}$C for 45 minutes. After the recovery period, cells were precipitated at 10,000 x g for 30 seconds, the majority of LB medium was discarded and precipitated cells were gently resuspended in 200μl of the remaining LB medium. Finally, the mixture was plated onto LB medium (LB plus 15gr bacteriological agar diluted in 1L $dH_2O$, autoclaved at 121$^{o}$C, for 20 min) agar plates containing appropriate antibiotic concentrations for each construct and incubated at 37$^{o}$C overnight. Optimisedprotocol included the following modifications. First, 10μl of each digestion-ligation reaction were mixed with *E. coli* DH5a chemo-competent cells. Second, recovery period was increased from 45 minutes to 90 minutes. Finally, transformed cells were plated onto LB medium plates with kanamycin 25μg/ml. For blue-white selection, precipitated cells were gently resuspended in 200μl of the remaining LB medium with 10μl of IPTG (0.5M) and 50μl of X-gal (40mg/ml).

## 2. Golden Gate-based plasmid construction

During the assembly of a CRISPR/Cas9 system optimisedfor gene editing in *L. japonicus*, a Golden Gate Modular cloning toolbox for plants was utilized consisting of two kits: the Golden Gate MoClo Plant Tool Kit and the Golden Gate MoClo Plant Parts Kit (Engler et al., 2014). Many acceptor vectors were used from the first kit as backbones to assemble level -1, level 0, level 1 and level 2 constructs. Contrarily, the second kit was used as a source of already domesticated level 0 modules or to generate new level 0 and level 1 modules. Golden Gate utilizes two type IIS restriction enzymes, BsaI and BbsI. An additional type IIS restriction enzyme was recruited, SapI, for the assembly of any sgRNA (also referred to as gRNA) cassette. During the assembly of the CRISPR/Cas9 level 0 and level 1 collection, a **short** digestion-ligation protocol had been used, while a **longer** cloning protocol ensured optimal assembly of higher-level multigene constructs (level 2 vectors) (Table 1, 2). For all the cloning reactions a 2:1 insert to vector ratio was employed.

**Table 1. Short digestion-ligation protocol (level -1, 0, 1 vectors)**

| Cycling conditions | | | |
|---|---|---|---|
| 37 $^{o}$C | 2 min | Repeat 50 times | BsaI/BbsI/SapI digestion |
| 16 $^{o}$C | 5 min | | T4 ligation |
| 80 $^{o}$C | 10-20min | Once | Enzyme inactivation |

**Table 2. Long digestion-ligation protocol (level 2 vectors)**

| Cycling conditions | | | |
|---|---|---|---|
| 37 $^{o}$C | 5 min | Repeat 60 times | BbsI digestion |
| 16 $^{o}$C | 5 min | | T4 ligation |
| 37 $^{o}$C | 10 min | Once | Acceptor linearization |
| 80 $^{o}$C | 10-20min | Once | Enzyme inactivation |

## 2.1. Level 0 modules

Construction of **level 0 Pro+5U promoter LjU6-1** (or **LjU6**) was performed by Magdalene Tsitsikle and is described in detail in her master thesis: "Mutagenesis of *Lotus japonicus* LSK-1 utilizing a CRISPR/Cas9 system". **Level 0 Pro+5U(f) promoter LjUbq1** was assembled in a Golden Braid acceptor by Nikolaos Delkis. Nevertheless, this module is still compatible for Golden Gate DNA assembly. Golden Gate MoClo Plant Parts Kit offers **level 0 CDS1 turboGFP** module, as well as the two utilized **level 0 3U+Ter terminators CaMV 35S** and **Act2**. **Level 0 gRNA scaffold** was also synthesised by Magdalene Tsitsikle and is described in her thesis. Finally, Dr. Constantine Garagounis constructed the **level 0 Pro+5U promoter LjUbq1**.

Primers LjUbpro_F and LjUbpro_R amplified **LjUbq1** promoter from plasmid pUB-GWS-Hyg via Q5® High-Fidelity DNA polymerase (New England Biolabs). Table 3 presents the sequence of each primer with underlined coloured nucleotides explaining the function of each region after amplification of target sequence. PCR reaction was purified via Macherey-Nagel™ NucleoSpin™ Gel and PCR Clean-up Kit, following manufacture's recommendations. Amplicon was mixed with level 0 acceptor Pro+5(U) pICH41295 of Golden Gate MoClo Plant Tool Kit in a BbsI/T4 ligase reaction. After transformation and blue-white selection, a white positive clone was identified and plasmid was sequenced (Sanger sequencing).

**Table 3. Primers for level 0 Pro+5U LjUbq1 module**

| Prime name | Primer sequence (5' → 3') | Amplicon size |
|---|---|---|
| LjUbpro_F | (TT GAAGAC AT GGAG) TGGAGAGAGGATTTTGAGGAA | 1122bp |
| LjUbpro_R | (TT GAAGAC AA CATT) CTGTAATCACATCAACAACAGA | |

Nucleotides within parentheses represent 5' primer overhangs. Each underlined and/or coloured region highlights its function after PCR amplification, when each 5' primer overhang has been turned into dsDNA. Nucleotides underlined and coloured with dark purple represent type IIS BbsI recognition site, while 4-nt overhangs, generated after digestion with BbsI, are underlined and coloured with orange. Underlined region represents two random nucleotides present between recogniton and cleavage site. Nucleotides coloured with grey at the 5' of each primer are added before the recognition site to allow better binding of the BbsI to the amplicon during digestion-ligation reaction. Nucleotides annealing to LjUbq1 promoter are depicted as non-underlined or coloured.

**Level 0 CDS1 Cas9 module**

PCR amplification of fragments and level -1 digestion-ligation reaction

Six primer sets were designed to amplify Cas9 with C' SV40 NLS in six fragments from plasmid pK7WGF2::hCas9 (gifted to us by Dr. Athanasios Dalakouras). Magdalene Tsitsikle designed all six sets of primers with Dr. Constantine Garagounis. Amplification was performed using Q5® High-Fidelity DNA polymerase (New England Biolabs), following manufacture's instructions. Each PCR cycling program included an initial denaturation step at 98°C for 30 sec, followed by 33 cycles of 98°C for 5 sec, the corresponding primer annealing temperature for 15 sec and 72°C for 20–30 seconds/kb. Final step included extension at 72°C for 2 min (Table 4). After gel extraction and purification with Macherey-Nagel™ NucleoSpin™ Gel and PCR Clean-up Kit, Cas9 fragments 1, 2 and 3 were subcloned into level -1 acceptor pAGM1311. Likewise, Cas9 fragments 4, 5 and 6 were subcloned into level -1 acceptor pAGM1311. Both were BsaI/T4 ligase reactions (Table 5). After transformation, bacteria were plated onto kanamycin 50μg/ml LB medium plates and blue-white selection was applied.

**Table 4. Primers for Cas9 domestication**

| Primer name | Primer sequence (5' → 3') | Tm | Ta | Extension time | Amplicon size |
|---|---|---|---|---|---|
| Cas9_1F | TT GGTCTC A ACAT A) ATG GACAAGAAGTACTCC | 57 | 58 | 16 sec | 1448bp |
| Cas9_1R | TT GGTCTC A) GAGT GATGGTTTCTTCTGA | 58 | | | |
| Cas9_2F | TT GGTCTC A) ACTC CCTGGAACTTCGAG | 64 | | | 310bp |
| Cas9_2R | TT GGTCTC A) TTGA AATAGTCCTCTTTGAGC | 59 | | | |
| Cas9_3F | TT GGTCTC A) TCAA AAAGATTGAATGTTTCGA | 57 | | | 491bp |
| Cas9_3R | TT GGTCTC A ACAA) GATGTGCTCGTGTAGACTG | 62 | | | |
| Cas9_4F | TT GGTCTC A ACAT) CATCGCTAATCTTGCAGG | 60 | 61 | 16 sec | 756bp |
| Cas9_4R | TT GGTCTC A) CTGA AATCAGACCACCAGCTTAG | 63 | | | |
| Cas9_5F | TT GGTCTC A) TCAG AAAGGACTTTCAGTTT | 58 | 59 | 17 sec | 877bp |
| Cas9_5R | TT GGTCTC A) ATCT TCGGGTGACCCTTT | 63 | | | |
| Cas9_6F | TT GGTCTC) A AGAT AATGAGCAGAAGCA | 58 | 59 | 8 sec | 419bp |
| Cas9_6R | TT GGTCTC A ACAA AAGC) CACACCTTCCTCTTCTTC | 61 | | | |

Nucleotides within parentheses represent 5' primer overhangs. Each underlined and/or coloured region highlights its function after PCR amplification, when each 5' primer overhang has been turned into dsDNA. Nucleotides underlined and coloured with green represent type IIS BsaI recognition site, while level -1 4-nt overhangs, generated after digestion with BsaI, are underlined and coloured with orange. Underlined region represents two random nucleotides present between recogniton and cleavage site. Nucleotides coloured with grey at the 5' of each primer are added before the recognition site to allow better binding of the BsaI to the amplicon during digestion-ligation reaction. Nucleotides outside of the parentheses anneal to the Cas9 coding sequence. Underlined blue-coloured nucleotides represent fusion sites for downstream level 0 cloning. Single dark red-coloured nucleotides introduce mismatches within the internal type IIS recognition sites. Tm and Ta values were calculated based on the online tool (https://tmcalculator.neb.com/#!/main) for use in Q5 High-Fidelity 2X Master Mix reaction. Extension time was calculated based on the manufacture's instructions for the same kit

**Table 5. Level -1 Cas9 digestion-ligation reaction**

| Cas9 frag 1-3 | BsaI/T4 dig-lig | Cas9 frag 4-6 | BsaI/T4 dig-lig |
|---|---|---|---|
| Vector pAMG1311(2611bp) | 50 ng | Vector pAMG1311(2611bp) | 50 ng |
| Cas9 frag1 | 54 ng | Cas9 frag4 | 7ng |
| Cas9 frag2 | 11 ng | Cas9 frag5 | 8 ng |
| Cas9 frag3 | 18 ng | Cas9 frag6 | 4 ng |
| BsaI 20U/µl | 0.5 µl | BsaI 20U/µl | 0.5 µl |
| T4 DNA Ligase 400U/µl | 0.25 µl | T4 DNA Ligase 400U/µl | 0.25 µl |
| Cutsmart 10x | 1 µl | Cutsmart 10X | 1 µl |
| T4 ligase buffer 10x | 1 µl | T4 ligase buffer 10x | 1 µl |
| $ddH_20$ | To 10 µl | $ddH_20$ | To 10 µl |

Enzymes and buffers are from New England Biolabs.

Diagnostic digestion

Plasmids from one Cas9 frag 1-3 and one Cas9 frag 4-6 white colony were isolated using Macherey Nagel NucleoSpin Plasmid- Miniprep kit, following manufacture's instructions. 1μl of purified Cas9 frag 1-3 plasmid was double digested with 0.2μl of PstI (NEB) and 0.2μl of SacI (NEB) restriction enzymes in 1μl of Cutsmart (10X) buffer (NEB) at a final volume of 10μl at 37°C for 2 hours. 1μl of purified Cas9 frag 4-6 plasmid was double digested with 0.2μl of BstXI (NEB) and 0.2μl of XhoI (NEB) restriction enzymes in 1μl of NEBuffer 3.1 (10X) buffer (NEB) at a final volume of 10μl at 37°C for 2 hours. Restriction digestions and 1μl of each uncut plasmid were analyzed over agaroze gel (1%) electrophoresis.

Level 0 digestion-ligation reaction

Level -1 Cas9 frag 1-3 and level -1 Cas9 frag 4-6 were subcloned into level 0 acceptor CDS1 pICH41308 of Golden Gate MoClo Plant Tool Kit through a BbsI/T4 ligase reaction (Table 6). After transformation, bacteria were plated onto spectinomycin 100μg/ml LB medium plates and blue-white selection was applied.

**Table 6. Level 0 CDS1 Cas9 digestion-ligation reaction**

| Component | BbsI/T4 dig-lig |
|---|---|
| Vector pICH41308 (2845bp) | 50 ng |
| Level -1 Cas9 1-3 (2169bp) | 19 ng |
| Level -1 Cas9 4-6 (1972bp) | 17 ng |
| BbsI 10U/μl (NEB) | 0.5 μl |
| T4 DNA Ligase 400U/μl (NEB) | 0.25 μl |
| Cutsmart 10x (NEB) | 1 μl |
| T4 ligase buffer 10x (NEB) | 1 μl |
| $ddH_20$ | To 10 μl |

Diagnostic digestion

Plasmid from a level 0 CDS1 Cas9 white colony was isolated using Macherey Nagel NucleoSpin Plasmid- Miniprep kit, following manufacture's instructions. 1μl of purified plasmid was double digested with 0.2μl of XhoI (NEB) and 0.2μl of SacI (NEB) restriction enzymes in 1μl of Cutsmart (10X) buffer (NEB) at a final volume of 10μl at 37°C for 2 hours. Restriction digestion and 1μl of uncut plasmid were analyzed over agaroze gel (1%) electrophoresis.

**Level 0 CDS1 Bar module**

PCR amplification of fragment and level 0 digestion-ligation reaction

Bar coding sequence was amplified from plasmid pICSL70005, using primers Bar_F and Bar_R in a Q5® High-Fidelity 2X mix reaction (Table 7). PCR cycling program included an initial denaturation step at 98°C for 5 min, followed by 30 cycles of 98°C for 5 sec, 59 °C for 15 sec and 72°C for 20 sec. Final step included extension at 72°C for 2 min. After gel extraction and purification via Macherey-Nagel™ NucleoSpin™ Gel and PCR Clean-up Kit, Bar fragment was subcloned into level 0 acceptor CDS1 pICH41308 of Golden Gate MoClo Plant Tool Kit through a BbsI/T4 ligase reaction (Table 8). Transformed bacterial cells were plated onto spectinomycin 100μg/ml LB medium plates for blue-white selection.

**Table 7. Primers for level 0 CDS1 Bar module**

| Prime name | Primer sequence (5' → 3') | Amplicon size |
|---|---|---|
| Bar_F | (TT GAAGAC AA G) ATG TCTCCTGAAAGAAGG | 579bp |
| Bar_R | (TT GAAGAC AA AA) GC TCAAATCTCAGTCACAGG | |

Nucleotides within parentheses represent 5' primer overhangs. Each underlined and/or coloured region highlights its function after PCR amplification, when each 5' primer overhang has been turned into dsDNA.

Nucleotides underlined and coloured with dark purple represent type IIS BbsI recognition site, while 4-nt overhangs, generated after digestion with BbsI, are underlined and coloured with orange. Underlined region represents two random nucleotides present between recogniton and cleavage site. Nucleotides coloured with grey at the 5' of each primer are added before the recognition site to allow better binding of the BbsI to the amplicon during digestion-ligation reaction. Nucleotides outside of the parentheses anneal to the Bar coding sequence.

**Table 8. Level 0 CDS1Bar digestion-ligation reaction**

| Component | BbsI/T4 dig-lig |
|---|---|
| Vector pICH41308 (2845bp) | 50 ng |
| Bar amplicon | 19,5 ng |
| BbsI 10U/µl (NEB) | 0.5 µl |
| T4 DNA Ligase 400U/µl (NEB) | 0.25 µl |
| Cutsmart 10x (NEB) | 1 µl |
| T4 ligase buffer 10x (NEB) | 1 µl |
| $ddH_20$ | To 10 µl |

Diagnostic digestion

Plasmid from a level 0 CDS1 Bar white colony was isolated using Macherey Nagel NucleoSpin Plasmid-Miniprep kit, following manufacture's instructions. 1µl of purified plasmid was digested with 0.2µl of StuI (NEB) restriction enzyme in 1µl of Cutsmart (10X) buffer (NEB) at a final volume of 10µl at 37°C for 2 hours. Restriction digestion and 1µl of uncut plasmid were analyzed over agaroze gel (1%) electrophoresis.

## 2.2. N' SV40 NLS ds oligo

N' SV40 NLS was synthesised by annealing two oligonucleotides, N'SV40NLS_F and N'SV40NLS_R (Table 9). For the hybridization reaction, 9µl of the forward primer (10pmol/µl) and 9 µl of the reverse primer (10pmol/µl) were added to 2µl of T4 ligase buffer (10X). Hybridization reactions were incubated as follows:

1. Incubation at 95◦C for 5 minutes to break secondary structures.
2. Incubation at room temperature for 20 minutes to allow the development of hydrogen bonds between complementary sequences.
3. Step one and two were repeated twice.
4. Hold on ice or at -20◦C.

**Table 9. Primers for N' SV40 NLS ds oligo**

| Prime name | Primer sequence (5' → 3') | ds oligo size |
|---|---|---|
| N'SV40NLS_F | CCAT ATGCCGAAGAAGAAGCGGAAGGTCGGAGCAGGAGG | 35bp |
| N'SV40NLS_R | CATT CCTCCTGCTCCGACCTTCCGCTTCTTCTTCGGCAT | |

Nucleotides underlined and coloured with orange are overhangs after oligonucletide annealing. They are complimentary to the right and left fusion site of level 0 Pro+5(U)f LjUbq1 and level 0 CDS1 Cas9 module, respectively.

## 2.3. Level 1 transcriptional units

### Level 1 reporter cassette

Level 1 digestion-ligation reaction

Level 0 modules LjUbq1 promoter (Pro+5U), turboGFP coding sequence and terminator of actin 2 gene were subloned into level 1 acceptor pICH41337 1F of 1st position and forward orientation through a BsaI/T4 ligase reaction (Table 10). Transformed bacterial cells were plated onto carbenicillin 100µg/ml LB medium plates for blue-white selection.

**Table 10. Level 1 reporter cassette digestion-ligation reaction**

| Level 1 reporter cassette 1F | BsaI/T4 dig-lig |
|---|---|
| Vector pICH47732 (4968bp) | 50 ng |
| LjUbq1 Pro+5U (1127bp) | 22ng |
| TurboGFP (714bp) | 14 ng |
| Act2 Ter (485bp) | 10 ng |
| BsaI 20U/μl | 0.5 μl |
| T4 DNA Ligase 400U/μl | 0.25 μl |
| Cutsmart 10x | 1 μl |
| T4 ligase buffer 10x | 1 μl |
| $ddH_20$ | To 10 μl |

Diagnostic digestion

Plasmid from a level 1 reporter cassette white colony was isolated using Macherey Nagel NucleoSpin Plasmid- Miniprep kit, following manufacture's instructions. 1μl of purified plasmid was double digested with 0.2μl of SpeI (NEB) and 0.2μl of BsaXI (NEB) restriction enzymes in 1μl of Cutsmart (10X) buffer (NEB) at a final volume of 10μl at 37°C for 2 hours. Restriction digestion and 1μl of uncut plasmid were analyzed over agaroze gel (1%) electrophoresis.

**Level 1 Cas9 expression cassette**

Level 1 digestion-ligation reaction

Level 0 modules LjUbq1 (Pro+5U(f)), N' SV40 NLS, Cas9 coding sequence and CaMV 35S terminator were subcloned into level 1 acceptor pICH47811 of 2nd position and reverse orientation through a BsaI/T4 ligase reaction (Table 11). Transformed bacterial cells were plated onto carbenicillin 100μg/ml LB medium plates for blue-white selection.

**Table 11. Level 1 Cas9 cassette digestion-ligation reaction**

| Level 1 Cas9 expression cassette 2R | BsaI/T4 dig-lig |
|---|---|
| Vector pICH4781 (4968bp) | 50 ng |
| LjUbq1 Pro+5U(f) (1126bp) | 23 ng |
| N' NLS | 0.5μl |
| Cas9 (4141bp) | 83 ng |
| 35S Ter (496bp) | 10 ng |
| BsaI 20U/μl | 0.5 μl |
| T4 DNA Ligase 400U/μl | 0.25 μl |
| Cutsmart 10x | 1 μl |
| T4 ligase buffer 10x | 1 μl |
| $ddH_20$ | To 10 μl |

Diagnostic digestion

Plasmid from a level 1 Cas9 expression cassette white colony was isolated using Macherey Nagel NucleoSpin Plasmid- Miniprep kit, following manufacture's instructions. 1μl of purified plasmid was double digested with 0.2μl of EcoRI (NEB) and 0.2μl of XmaI restriction enzymes in 1μl of Cutsmart (10X) buffer (NEB) at a final volume of 10μl at 37°C for 2 hours. Restriction digestion and 1μl of uncut plasmid were analyzed over agaroze gel (1%) electrophoresis.

**Level 1 gRNA acceptor**

Level 1 digestion-ligation reaction

Level 0 modules LjU6 promoter and gRNA scaffold were subcloned into level 1 acceptors pICH47751 of 3rd

position and forward orientation and pICH47761 of 4th position and forward orientation through a BsaI/T4 ligase reaction (Table 12). Transformed bacterial cells were plated onto carbenicillin 100μg/ml LB medium plates for blue-white selection.

**Table 12. Level 1 gRNA acceptor digestion-ligation reaction**

| gRNA acceptor 3F/4F | BsaI/T4 dig-lig |
|---|---|
| Vector pICH47751/ pICH477611 (4968bp) | 50 ng |
| LjU6 Pro+5U (812bp) | 16 ng |
| gRNA scaffold (119bp) | 3 ng |
| BsaI 20U/μl | 0.5 μl |
| T4 DNA Ligase 400U/μl | 0.25 μl |
| Cutsmart 10x | 1 μl |
| T4 ligase buffer 10x | 1 μl |
| $ddH_20$ | To 10 μl |

Diagnostic digestion

Plasmids from one level 1 gRNA acceptor 3F and one gRNA acceptor 4F white colony were isolated using Macherey Nagel NucleoSpin Plasmid- Miniprep kit, following manufacture's instructions. 1μl of each purified plasmid was digested with 0.2μl of PsiI (NEB) and 0.2μl of XmaI restriction enzymes in 1μl of Cutsmart (10X) buffer (NEB) at a final volume of 10μl at 37°C for 2 hours. Restriction digestion and 1μl of uncut plasmid were analyzed over agaroze gel (1%) electrophoresis.

**Level 1 selection marker**

Level 1 digestion-ligation reaction

Level 0 modules LjUbq1 promoter (Pro+5U), Bar coding sequence and CaMV 35S terminator were subcloned into level 1 acceptors pICH47831 of 4th position and reverse orientation and pICH47841 of 5th position and reverse orientation through a BsaI/T4 ligase reaction (Table 13). Transformed bacterial cells were plated onto carbenicillin 100μg/ml LB medium plates for blue-white selection.

**Table 13. Level 1 selection marker digestion-ligation reaction**

| Selection marker 4R/5R | BsaI/T4 dig-lig |
|---|---|
| Vector pICH47831/ pICH47841 (4968bp) | 50 ng |
| LjUbq1 Pro+5 (1126bp) | 23 ng |
| Bar CDS1 (553bp) | 11 ng |
| 35S Ter (496bp) | 10 ng |
| BsaI 20U/μl | 0.5 μl |
| T4 DNA Ligase 400U/μl | 0.25 μl |
| Cutsmart 10x | 1 μl |
| T4 ligase buffer 10x | 1 μl |
| $ddH_20$ | To 10 μl |

Diagnostic digestion

Plasmids from one level 1 selection marker 4R and one selection marker 5R white colony were isolated using Macherey Nagel NucleoSpin Plasmid- Miniprep kit, following manufacture's instructions. 1μl of each purified plasmid was digested with 0.2μl of BsaAI (NEB) and 0.2μl of XhoI restriction enzymes in 1μl of Cutsmart (10X) buffer (NEB) at a final volume of 10μl at 37°C for 2 hours. Restriction digestion and 1μl of uncut plasmid were analyzed over agaroze gel (1%) electrophoresis.

## 2.4. gRNA cassettes

### 2.4.1. Selection of target sites and design of gRNA sequences

Genomic and coding sequences of ***AMY2*** (Lj3g3v1983420.2, AF478455_1), ***CYP71D353*** (Lj3g3v1983600.1, KF460438.1) and ***LSK1*** **(**Lj0g3v0083059.1, AB113573.1) were accessed from Lotus Base (https://lotus.au.dk) (Mun, Bachmann, Gupta, Stougaard, & Andersen, 2016) and/or Genbank (https://www.ncbi.nlm.nih.gov/genbank/) and 5'UTRs, exonic and intronic regions were determined. Two web tools were used to search for putative CRISPR/Cas9 target sites, CRISPR-P 2.0 (Liu et al., 2017) and CRISPRdirect (Naito, Hino, Bono, & Ui-Tei, 2014).

After selecting *L. japonicus* (v3.0) as the reference genome, the U6 promoter to drive the gRNA expression, the gRNA scaffold from Streptococcus pyogenes system, 20-nt as the size of the gRNA sequence and 5' NGG 3' as the PAM sequence recognised by Cas9 endonuclease of Streptococcus pyogenes, genomic regions of *AMY2* and *CYP71D353* were submitted to CRISPR-P 2.0. gRNA sequences with four consecutive T's were excluded. CRISPR guide sequences with a minimum on-score of 0.6 were considered for further analysis. Any potential off-target sites harbored enough mismatches at crucial regions as to allow for a low off-score. Additionally, gRNA sequences with off-target sites in coding regions were avoided.

CRISPRdirect was used to search for target sites in all three genes. Genomic regions or just exons were submitted, using Japanese trefoil (*Lotus japonicus*) genome, build 3.0 as the reference genome and choosing 5' NGG 3' as the PAM sequence. GuideRNA sequences with four consecutive T's were again excluded, since it can cause Pol III termination. Exonic on-target sites with the lowest number of off-target sites were considered for further analysis.

For *AMY2*, five gRNA sequences were selected (AMY2-gRNA1-5), including two from previous analysis. Two gRNA sequences targeting *CYP71D353* (CYP71D353-gRNA1-2) were selected. Finally, Nikolaos Delkis selected four target sites for *LSK1* (LSK-gRNA1-4).

### 2.4.2. Annealing of oligonucleotides

Each gRNA sequence was synthesised by annealing two oligonucleotides (Table 14). 9µl of the forward primer (10pmol/µl) and 9 µl of the reverse primer (10pmol/µl) were added to 2µl of T4 ligase buffer (10X). Hybridization reactions were incubated as follows:

1. Incubation at 95◦C for 5 minutes.
2. Incubation at room temperature for 20 minutes.
3. Step one and two were repeated twice.
4. Hold on ice or at -20◦C.

**Table 14. gRNA oligonucleotides**

| Primer name | Primer sequence (5' → 3') | Tm | GC% | Targeted gene region |
|---|---|---|---|---|
| AMY2-gRNA1_ F | TTC G CCAACCATATATACCCACCG | 61 | 50 | AMY2-intron1 |
| AMY2-gRNA1_ R | AACCGGTGGGTATATATGGTTGG C | | | |
| AMY2-gRNA2_ F | TTC G TACATATGCATGCGAATACT | 57 | 38 | AMY2-exon4 |
| AMY2-gRNA2_ R | AACAGTATTCGCATGCATATGTA C | | | |
| AMY2-gRNA3_ F | TTC G CGACACACATAGCTTCTTGG | 61 | 50 | AMY2-exon4 |
| AMY2-gRNA3_ R | AACCCAAGAAGCTATGTGTGTCG C | | | |
| AMY2-gRNA4_ F | TTC G CACTGCATAGAAAAGCCATA | 58 | 42 | AMY2-intron9 |
| AMY2-gRNA4_ R | AACTATGGCTTTTCTATGCAGTG C | | | |
| AMY2-gRNA5_F | TTC G TGACATCTCGAGTTTCACAT | 59 | 42 | AMY2-intron10 |
| AMY2-gRNA5_R | AACATGTGAAACTCGAGATGTCA C | | | |
| CYP71D353-gRNA1_ F | TTC G TCTTGGATAACTCTCTCAGG | 61 | 46 | CYP71D353-exon 1 |
| CYP71D353-gRNA1_ R | AAC CCTGAGAGAGTTATCCAAGA C | | | |
| CYP71D353-gRNA2_ F | TTC G TAATCACAAGTAGCCAGCGT | 65 | 46 | CYP71D353-exon 1 |
| CYP71D353-gRNA2_ R | AAC ACGCTGGCTACTTGTGATTA C | | | |
| LSK1-gRNA1_ F | TTC G TCGAGACGGACAGCCGAAGC | 67 | 63 | LSK1-exon3 |
| LSK1-gRNA1_ R | AACGCTTCGGCTGTCCGTCTCGA C | | | |
| LSK1-gRNA2_ F | TTC GTGGTTGGGACTGGATCCTT | 62 | 52 | LSK1-exon4 |
| LSK1-gRNA2_ R | AACAAGGATCCAGTCCCAACCAC | | | |
| LSK1-gRNA3_ F | TTC GGTCAACTGTGCGCATAACC | 63 | 52 | LSK1-exon5 |
| LSK1-gRNA3_ R | AACGGTTATGCGCACAGTTGACC | | | |
| LSK1-gRNA4_ F | TTC G CTCAGATTAAAGCGCACCCG | | | |

| | | | | |
|---|---|---|---|---|
| LSK1-gRNA4_ F | TTC **G** CTCAGATTAAAGCGCACCCG | 64 | 54 | LSK1-exon10 |
| LSK1-gRNA4_ R | AACCGGGTGCGCTTTAATCTGAG **C** | | | |

Nucleotides coloured with orange represent 3-nt overhangs that are complimentary to gRNA acceptor fusion sites after digestion with SapI. For each gRNA sequence that does not start with guanine, it was added at the beginning, after the 5' overhang, during primer synthesis. It is coloured with light blue. This is why some oligonucleotides have 24 and others 23 nucleotides.

### 2.4.3. Cloning of dsoligos into gRNA acceptors

Each ds oligo gRNA was subcloned into the appropriate gRNA acceptor. gRNAs were mixed with gRNA acceptor 3F or gRNA acceptor 4F in a SapI/T4 ligase reaction (Table 15). Table 16 shows the list of generates gRNA cassettes.

**Table 15. gRNA cassette digestion-ligation reaction**

| gRNA cassette | SapI/T4 dig-lig |
|---|---|
| gRNA acceptor 3F/4F (5313bp) | 50 ng |
| Ds oligo gRNA (23 or 24 bp) | 1µl |
| SapI 10U/µl | 1 µl |
| T4 DNA Ligase 400U/µl | 0.25 µl |
| Cutsmart 10x | 1 µl |
| T4 ligase buffer 10x | 1 µl |
| $ddH_20$ | To 10 µl |

**Table 16. List of gRNA cassettes**

| Ds oligo gRNA | gRNA acceptor |
|---|---|
| AMY2-gRNA1_ F/R | gRNA acceptor 4F |
| AMY2-gRNA2_ F/R | gRNA acceptor 3F |
| AMY2-gRNA3_ F/R | gRNA acceptor 3F |
| AMY2-gRNA4_ F/R | gRNA acceptor 3F |
| AMY2-gRNA5_ F/R | gRNA acceptor 4F |
| CYP71D353-gRNA1_ F/R | gRNA acceptor 3F |
| CYP71D353-gRNA2_ F/R | gRNA acceptor 4F |
| LSK1-gRNA1_ F/R | gRNA acceptor 3F |
| LSK1-gRNA2_ F/R | gRNA acceptor 3F |
| LSK1-gRNA3_ F/R | gRNA acceptor 4F |
| LSK1-gRNA4_ F/R | gRNA acceptor 4F |

5µl of digestion-ligation reaction were transformed into DH5a chemo-competent cells and transformed bacteria were plated onto carbenicillin 100µg/ml plates. Plasmids were isolated using Macherey Nagel NucleoSpin Plasmid- Miniprep kit, following manufacture's instructions.

## 2.5. Level 2 vectors

Level 1 transcriptional unit reporter cassette, Cas9 expression cassette, gRNA cassette(s) and plant selection marker were assembled into level 2 vector pAGM4723 in a BbsI/T4 ligase reaction. The appropriate end-linker was used to connect the far right selection marker to the right overhang of pAGM4723 (Table 17). A vector containing everything but the 20nt sequence of gRNA(s) was also assembled as the negative control (Table 18). 10µl of each reaction was transformed into DH5a chemo-competent cells and transformed cells were plated onto kanamycin 25µg/ml LB medium plates

**Table 17. Level 2 multigene vector digestion-ligation reaction**

| Level 2 vector with 2 gRNA cassettes | BbsI/T4 dig-lig | Level 2 vector with 1 gRNA cassettes | BbsI/T4 dig-lig |
|---|---|---|---|
| Vector pAGM4723 (12919bp) | 50 ng | Vector pAGM4723 (12919bp) | 50 ng |
| Reporter cassette **1F** (2351bp) | 50 ng | Reporter cassette **1F** | 50 ng |
| Cas9 expression cassette **2R** (5593) | 114 ng | Cas9 expression cassette **2R** | 114 ng |
| gRNA cassette **3F** (987bp) | 20 ng | gRNA cassette **3F** | 20 ng |
| gRNA cassette **4F** (987bp) | 20 ng | - | - |
| Plant selection marker **5R** (1908bp) | 40 ng | Plant selection marker **4R** | 40 ng |
| End-link pICH41800 (**EL5**) (24bp) | 1 ng | End-link pICH41780 (**EL4**) | 1 ng |
| BbsI 10U/µl | 0.5 µl | BbsI 10U/µl | 0.5 µl |
| T4 DNA Ligase 400U/µl | 0.25 µl | T4 DNA Ligase 400U/µl | 0.25 µl |
| Cutsmart 10x | 1 µl | Cutsmart 10x | 1 µl |
| T4 ligase buffer 10x | 1 µl | T4 ligase buffer 10x | 1 µl |
| $ddH_20$ | To 10 µl | $ddH_20$ | To 10 µl |

Diagnostic digestion

White colonies were selected and plasmids were isolated using Macherey Nagel NucleoSpin Plasmid-Miniprep kit, following manufacture's instructions. 0.8µl of each purified plasmid was digested with 0.2µl of XmnI (NEB) restriction enzyme in 1µl of Cutsmart (10X) buffer (NEB) at a final volume of 10µl at 37$^{o}$C for 2 hours. Restriction digestions and 1µl of uncut plasmid were analyzed over agaroze gel (1%) electrophoresis.

**Table 18. Level 2 multigene vector Cas9-only digestion-ligation reaction**

| Level 2 vector Cas9-only (negative control) | BbsI/T4 dig-lig |
|---|---|
| Vector pAGM4723 (12919bp) | 50 ng |
| Reporter cassette **1F** (2351bp) | 50 ng |
| Cas9 expression cassette **2R** (5593) | 114 ng |
| gRNA acceptor **3F** (987bp) | 20 ng |
| gRNA acceptor **4F** (987bp) | 20 ng |
| Plant selection marker **5R** (1908bp) | 40 ng |
| End-link pICH41800 (**EL5**) (24bp) | 1 ng |
| BbsI 10U/µl | 0.5 µl |
| T4 DNA Ligase 400U/µl | 0.25 µl |
| Cutsmart 10x | 1 µl |
| T4 ligase buffer 10x | 1 µl |
| $ddH_20$ | To 10 µl |

Table 19 presents the complete list of assembled CRISPR/Cas9 binary vectors. All binary vectors harbor two T-DNA repeat regions, left and right border, defining the T-DNA region. T-DNA region, containing all level 1 transcriptional units, will be transferred from the level 2 vector to the plant genome through *A. rhizogenes* LBA 1334.

**Table 19. List of assembled CRISPR/Cas9 binary vectors**

| Target gene region | CRISPR/Cas9 binary vectors |
|---|---|
| AMY2: intron 1 and exon 4 | AMY2:gRNA1-gRNA2 |
| AMY2: exon 4 | AMY2:gRNA2 |
| AMY2: exon 4 | AMY2:gRNA3 |

| | |
|---|---|
| AMY2: intron 9 and intron 10 | AMY2:gRNA4-gRNA5 |
| CYP71D353: exon 1 | CYP71D353:gRNA1-gRNA2 |
| LSK1: exon 3 and exon 5 | LSK1:gRNA1-gRNA3 |
| LSK1: exon 4 and exon 10 | LSK1:gRNA2-gRNA4 |
| - | Cas9-only (negative control) |

## 3. *Agrobacterium rhizogenes*-mediated hairy root transformation

### 3.1. *L. japonicus* seed scarification/sterilization and germination

*Lotus japonicus* seeds (cultivars MG20 and Gifu) were submerged in sulfuric acid ($H_2SO_4$) and incubated for 15-20 min at room temperature by gentle shaking. Acid was carefully discarded and seeds were washed five times with sterilized cold $dH_2O$. Subsequently, seeds were sterilized with a solution (20% bleach-0.1% tween20) by gentle shaking for 12-15 min, or until they turned white, at room temperature. Seeds were washed eight times, or until the smell of bleach disappeared, with sterilized $ddH_2O$. Finally, scarified and sterilized seeds were submerged in sterilized $ddH_2O$ and maintained in the dark at 4$^o$C for 2 days. After two days, swollen seeds were transferred onto ½ MS agar (0.8%) petri dishes and covered with aluminium foil. Plates were placed vertically and maintained at **22$^o$C** for 3 days. After three days, aluminium foil was removed and percentage of germination was evaluated. Subsequently, germinated seeds were maintained in a 16h light/ 8h dark cycle at 22$^o$C for four days.

### 3.2. *Agrobacterium rhizogenes* transformation and culture preparation

CRISPR/Cas9 binary vectors were introduced into *Agrobacterium rhizogenes* LBA1334 strain by electroporetion. 1μl of each vector (50ng/μl) was added to 100μl of *A. rhizogenes* electro-competent cells. After gentle mixing, the mixture was transferred to a sterilized cold cuvette without introducing air bubbles (by gentle tapping of cuvette on the bench). The cuvette was then inserted into the cuvette chamber of the micropulser electroporetor (Biorad). After the electoporetion pulse was delivered, 1 ml of liquid LB was added and cells were gently resuspended. Mixture was then transferred to the original tube and incubated at 28$^o$C for 3-4 hours. After recovery period, cells were precipitated at 10,000 x g for 30 seconds and the majority of LB medium was discarded. Cells were resuspended in the remaining 200μl of LB medium and plated onto LB medium agar plates containing either kanamycin 25μg/ml or 50μg/ml. Colonies appeared after incubation of plates at 28$^o$C for 2 days.

Five days after the scarification process, at least one liquid *A. rhizogenes* culture of 5ml harboring each construct was prepared from a single colony. All liquid cultures were grown at 28$^o$C (optimal temperature) for 2 days supplemented with either kanamycin 25μg/ml or 50μg/ml. Afterwards, 0.5 or 1ml of each culture were transferred to LB plates with either kanamycin 25μg/ml or 50μg/ml and grown at 28$^o$C for another 2 days. At the day of the hairy root transformation each plate contained a thick lawn of bacterial colonies.

### 3.3. Hairy root transformation

Each bacterial culture had to be first diluted, by adding 1-3 ml of sterilized $ddH_2O$ and suspending the bacterial colonies by scrapping the surface. A small amount of the resuspended culture was transferred to sterilized filter paper. Four-day-old seedlings (as day one is considered the day aluminium foil was removed) were placed on the wet filter paper and the primary root of each seedling was cut right under the hypocotyl with a sterile blade. A small amount of resuspended bacterial culture was injected into the wounded site. Seedlings were then transferred to ½ MS (1/2X Murashige and Skoog salt, 2%w/v sucrose, 0.8% bacteriological agar, pH=5.5 with KOH, autoclaved at 121$^o$C for 20min)-agar square plates (15 seedlings per plate). Plates were covered with aluminium foil, placed vertically and maintained in a 16h light/8h dark cycle at 22$^o$C for 2 days. After two days, aluminium foil was removed and plates were maintained in the same conditions for another 2 days. Five days after the co-cultivation with *A. rhizogenes,* explants were transferred to fresh ½ MS-agar square plates with 300μg/ml ceftazidime (10 explants per plate). Plates were again placed vertically and held in a 16h light/8h dark cycle at 22$^o$C. If neccessary, plants with emerging hairy roots were retransferred to fresh ½ MS-agar square plates with 300μg/ml ceftazidime.

Six different bacterial cultures were used with five of them harbouring CRISPR/Cas9 binary vectors targeting a specific gene and one carrying Cas9 expression cassette only (all level one cassettes were present,

except for the 20nt sequence of gRNA). Seedlings infected with the last culture were the control lines (negative control). Two *L. japonicus* cultivars were used, MG20 and Gifu. Table 20 displays the target genes, which CRISPR/Cas9 vectors were used to target each gene, including Cas9-only vector, number and cultivar of seedlings co-cultivated with each culture and number of sampled plants for each construct.

**Table 20. Hairy root transformation**

| Target gene | *A.rhizogenes* culture harbouring CRISPR/Cas9 binary vector | # of seedlings used in each root transformation | # of sampled plants |
|---|---|---|---|
| AMY2 | gRNA1-gRNA2 | 50 MG20 | 40 |
| | gRNA4-gRNA5 | 80 MG20 | 60 |
| CYP71D353 | gRNA1-gRNA2 | 50 MG20 | 40 |
| LSK1 | gRNA1-gRNA3 | 50 Gifu | 40 |
| | gRNA2-gRNA4 | 50 MG20 | 40 |
| Control lines | Cas9-only (negative control) | 20 MG20 | 10 |
| | Cas9-only (negative control) | 20 Gifu | 10 |

## 3.4. Identification and sampling of GFP positive transgenic hairy roots

Twenty seven-twenty eight days since the beginning of the experiment, each hairy root emerging from the hypocotyl was individually sampled and stored at -80$^{o}$C for analysis. Distinction between transformed and non-transformed roots occurred during this stage, instead of searching for T-DNA positive samples later by performing diagnostic PCR on all root samples. Since each T-DNA carried a GFP reporter cassette close to the left border, roots exhibiting green fluorescence under a microscope would be characterized as transformed. During sampling, plants were studied under the GFP fluorescence filter of Leica StereomicroscopeMZ 10 F and GFP-positive roots were stored.

## 3.5. Genomic extraction from root samples using CTAB protocol

Root samples were firstly pulverized using the TissueLyser II system (Qiagen). Each tube, containing one big or two small stainless steel beads, was submerged in liquid nitrogen and then shaken long enough as to allow the roots to turn into powder. Next, 300μl of heated CTAB extraction buffer, containing 1μl of b-mercaoptoethanol, were added to each tube. Mixtures was mixed thoroughly with a vortex and homogenates were transfered to a heat bath at 65$^{o}$C for 20 minutes. Following the incubation period, homogenates were centrifuged for 5 minutes at 11,000 x g. Supernatants were transferred to new tubes and 150μl of chloroform/isoamyl alcohol (24:1) were added to each tube. Mixtures were then mixed using vortex until turned into a white cloudy solution. After a centrifugation for 5 minutes at 11,000 x g, aqueous upper phase was carefully sucked and transferred to a new tube. Since DNA was diluted in this phase, precipitation was achieved by adding equal volume of cold isopropanol, followed by incubation at room temperature for 10 minutes and centrifugation for 15 minutes at 11,000 x g. Supernatant was discarded without disturbing the pellet. Subsequently, precipitated DNA was washed with 500μl of ice-cold 70% ethanol for 5 minutes at 11,000 x g. After carefully discarding ethanol, tubes were air-dried. When ethanol evaporated, DNA was dissolved in 15μl of sterilized dd$H_2O$ containing RNAse solution A (1μl of RNAse A diluted to 50μl of dd$H_2O$) by scrapping the bottom of the tube with the pipette tip. Quantity and quality of genomic samples was determined in Q3000 UV spectrophotometer (Quawell). Genomic samples were stored at -20$^{o}$C until further analysis.

## 3.6. Mutation analysis

### 3.6.1. PCR assay

Regions including target sites of gRNAs were amplified by PCR using Q5® High-Fidelity DNA polymerase (New England Biolabs). Amplification was performed in either 25μl or 50μl final volume (Table 21). PCR cycling program included an initial denaturation step at 98$^{o}$C for 30 sec, followed by 35 cycles of 98$^{o}$C for 5

sec, the corresponding primer annealing temperature for 15 sec and 72°C for 20–30 seconds/kb. Final step included extension at 72°C for 2 min. Table 22 displays target gene regions, primer sets used to amplify them and primer sequences with melting temperature and annealing temperature values for each set. Furthermore, amplicon size from wild type allele and extension time are included. 3μl of each reaction were analyzed by agarose-gel 1% electrophoresis.

**Table 21. Q5® High-Fidelity DNA polymerase reaction for mutation analysis**

| Component | 25 μl Reaction | 50 μl Reaction |
|---|---|---|
| Q5 High-Fidelity 2X Master Mix | 12.5 μl | 25 μl |
| 10 μM Forward Primer | 1.25 μl | 2.5 μl |
| 10 μM Reverse Primer | 1.25 μl | 2.5 μl |
| Template DNA | 1μl (1-200ng/μl) | 1μl (1-200ng/μl) |
| $ddH_2O$ | to 25 μl | to 50 μl |

**Table 22. PCR parameters for mutation analysis**

| Target gene region | Primers | Primer sequence (5' → 3') | Tm | Ta | Extension time | Amplicon size |
|---|---|---|---|---|---|---|
| AMY2 intron1 | AMY2-intron1_F | ATTCAAACCATACTTCGCAC | 60 | 61 | 17 sec | 566bp |
| | AMY2-intron1_R | CTTGGCGTATTCTCTCTTCG | 63 | | | |
| AMY2 exon4 | AMY2-exon4_F | CGGATGGAATTGTTATCCAAG | 61 | 61 | 17 sec | 506bp |
| | AMY2-exon4_R | CTGGATGCATAGGAAGAAATG | 60 | | | |
| AMY2 exon10 | AMY2-exon10_F | TCAAGAATTCTCAGGTCTGCTT | 63 | 55* | 19 sec | 614bp |
| | AMY2-exon10_R | GCACAGCACAAAAAAAGAATT | 60 | | | |
| CYP71D353 Exon 1 | CYP71D-exon1_F | GATGGAGGATCATTTCTTATCC | 60 | 61 | 19 sec | 851bp |
| | CYP71D-exon1_R | TGATGGTCAAATGAAATTCAGG | 60 | | | |

Each primer set was used to amplify the corresponding gene region including the target site(s). Tm and Ta values were calculated based on the online tool (https://tmcalculator.neb.com/#!/main) for use in Q5 High-Fidelity 2X Master Mix. Extension time was calculated based on the manufacture's instructions for the same kit. *For primer set AMY2-exon10_F and AMY2-exon10_R annealing temperature was decreased from recommended 61°C to 55°C, because some samples exhibited a faint band during gel electrophoresis.

### 3.6.2. T7 endonuclease I assay

T7 endonuclease I assay was used to assess gene editing in PCR products. PCR reactions were purified via Macherey-Nagel™ NucleoSpin™ Gel and PCR Clean-up Kit, following manufacture's recommendations. PCR products were eluted in 20-30 μl of water. Quantity of eluted PCR products was determined in Q3000 UV spectrophotometer (Quawell). PCR productes were stored at -20°C until further analysis. T7 endonuclease I assay reactions were assembled in PCR tubes as follows**.**

**Table 23. Annealing reaction**

| Component | Annealing reaction |
|---|---|
| PCR product | 250ng |
| 10X NEBuffer 2 | 2μl |
| $ddH_2O$ | To 19μl |

PCR products were denaturated and annealed in a thermocycler with the following conditions.

**Table 24. Denaturation and reannealing conditions**

| Step | Temperature | Ramp rate | Time |
|---|---|---|---|
| Initial Denaturation | 95°C | | 5 min |
| Annealing | 95-85°C | -2°C/sec | |
| | 85-25°C | -0.1°C/sec | |
| Hold | 4°C | | Hold until T7 I digestion |

0.8μl of T7 endonuclease I was added to each tube. Reactions were incubated at 37°C for 1 hour. 0.5μl of 0.5M EDTA (pH=8) was added to stop the reaction. To assess the mutation efficacy of each gRNA, T7 endonuclease I assay reactions were subjected to agarose gel (1.3%) electrophoresis.

# Results

## 1. Assembly of a Golden Gate cloning toolkit optimised for gene editing in *L. japonicus*

Our aim was to develop a CRISPR/Cas9-based genome editing system, optimised for mutagenesis in *L. japonicus*, for use in functional analyses. We needed a cloning approach to rapidly generate any CRISPR/Cas9 binary vector. Therefore, we adopted Golden Gate DNA assembly method, a type IIS restriction enzyme-based cloning approach (Weber et al., 2011). Type IIS restriction endonucleases differ from classical type II enzymes, as they cleave outside of their recognition site, without any sequence requirement for the identity of the bases at the cleavage site. This particular characteristic allows for **one-pot, one-step assembly**. One pot refers to the fact that digestion of inserts and backbone, and ligation, can be done in the same tube. Furthermore, assembly of multiple parts in the same vector in one digestion-ligation reaction is achievable, justifying the one step notion. To create a CRISPR/Cas9 Golden Gate–based system optimised for mutagenesis in *L. japonicus,* we started by selecting or generating appropriate standard parts (level 0 modules in terms of Golden Gate cloning) for the **CRISPR/Cas9 level 0 collection**. Then, based on the order and orientation of each expression cassette in the final multigene construct (level 2 vector), we assembled selected level 0 modules into appropriate level 1 acceptor backbones to make complete transcriptional units for the **CRISPR/Cas9 level 1 collection**.

### 1.1. CRISPR/Cas9 level 0 collection of stardard parts

In Golden Gate cloning, the genetic grammar is broken down into standard parts, such as promoters, coding sequences and other regulatory elements. At the bottom of the hierarchy are the level 0 modules, which are level 0 vectors containing standard parts of the genetic grammar. Each type of level 0 module contains a pre-determined set of fusion sites that defines its position in the transcriptional unit (Fig. 1. Our plan was to create a Golden Gate-compatible collection of promoters, coding sequences, terminators and other necessary regulatory components that would facilitate assembly of CRISPR/Cas9 binary vectors. To create this **CRISPR/Cas9 level 0 collection**, we utilized the following vector kits: the Golden Gate MoClo Plant Tool Kit and the Golden Gate MoClo Plant Parts Kit (Engler et al., 2014). The first kit contains a set of vectors for constructing, among others, level 0, level 1 and level 2 vectors, while the second kit includes a variety of level 0 modules for use in plant biotechnology. Furthermore, we generated additional level 0 modules in order to conform CRISPR/Cas9 technology to the principles of Golden Gate cloning. All original and newly synthesised modules are required to: **i)** be flanked by two BsaI sites in opposing orientations, **ii)** be flanked by the same pair of fusion sites, for parts of the same type (e.g. different promoters share the same fusion sites), **iii)** be freed of internal recognition sites for type IIS enzymes BsaI and BbsI, and **iv)** be cloned into a spectinomycin resistant plasmid backbone.

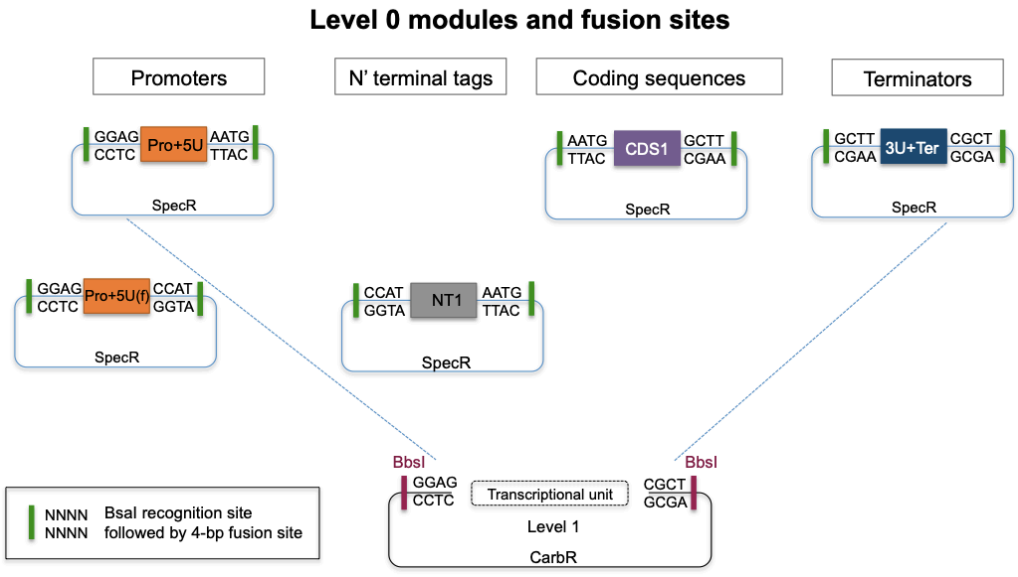


**Figure 1. In Golden Gate, stardard parts of genetic grammar are cloned into level 0 modules with specific fusion sites**. Level 0 modules are assembled into level 1 transcriptional units with type IIS BsaI restriction enzyme. Level 1 transcriptional units are flanked by two type IIS BbsI recognition sites for downstream assembly into level 2 vectors.

## Level 0 promoters

A CRISPR/Cas9 system requires a single guideRNA (sgRNA) to guide the Cas9 endonuclease for site-specific genome editing. Cas9 is usually driven by endogenous, Polymerase II-dependent housekeeping genes' (i.e., ubiquitin gene) promoters, or a strong constitutive promoter like Cauliflower mosaic virus (CaMV) 35s promoter (Kumar & Jain, 2015). On the contrary, endogenous Polymerase III promoters, like U6 or U3, are efficient in transcribing high levels of small non-coding RNAs (Kumar & Jain, 2015). We decided on an endogenous ***L. japonicus*** **U6** promoter to drive the expression of sgRNAs (Wang et al., 2016, 2019). As for the Cas9 expression, because the transcriptional activity of the CaMV 35s is quite weak in *L. japonicus* tissues, we chose **LjUbq1**, the promoter of an endogenous ubiquitin gene fused to its first intron (Maekawa et al., 2008).

**LjUbq1** was PCR amplified from plasmid pUB-GWS-Hyg, using primers LjUbpro_F and LjUbpro_R in a Q5® High-Fidelity 2X mix reaction. The 5'overhangs of these primers introduced the following things on each end of the producing amplicon: a BbsI recognition site, followed by the appropriate 4-bp sequence, 5'GGAG3' on the left side and 5'AATG3' on the right side (Fig. 2). After PCR clean up, LjUbq1 amplicon was cloned into level 0 acceptor Pro+5(U) pICH41295 of Golden Gate MoClo Plant Tool Kit in a BbsI/T4 ligase reaction (Fig. 2). Cloning reaction was transformed into DH5a chemo-competent cells and transformed bacteria were plated onto spectinomycin 100μg/ml LB medium plates. Blue-white selection was applied and a level 0 LjUbq1 clone was verified through Sanger sequencing. Dr. Constantine Garagounis completed this level 0 module.

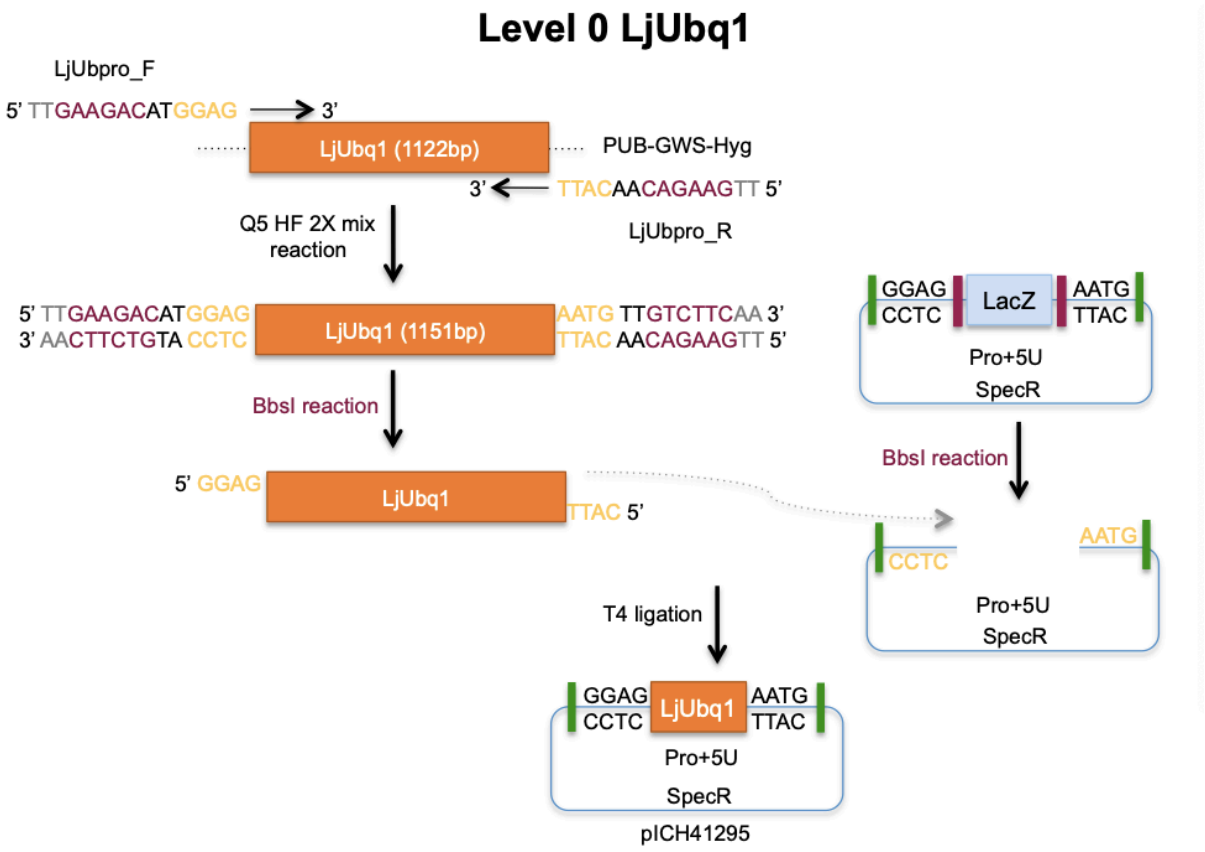



**Figure 2**. **Construction of level 0 Pro+5U LjUbq1 module.**

**LjUbq1** promoter was also cloned into pUPD2 vector of Golden Braid cloning toolkit. This level 0 module

of LjUbq1 has a different right fusion site. Instead of 5'AATG3', it is flanked by 5'CCAT3', which is complimentary to the left 5'CCAT3' site of an amino terminal level 0 module. Amplified LjUbq1 was cloned into pUPD2 through a BsmBI/T4 ligase cloning reaction, which follows the same principles as a BbsI/T4 ligase reaction. Nikolaos Delkis, a master student in our laboratory, performed and verified, through Sanger sequencing, this particular level 0 module.

Finally, **LjU6-1** (or **LjU6**) promoter was PCR amplified in three fragments from *L. japonicus* Gifu genomic DNA, in order to remove one internal BbsI site and one internal BsaI site. Two internal reverse primers introduced a single mismatch at each recognition site. Fragment one of LjU6 promoter and fragments two and three were subcloned into two separate level -1 universal acceptors pAGM1311 of Golden Gate MoClo Plant Tool Kit through a BsaI/T4 ligase reaction. Subsequently, both parts were released and subcloned into level 0 acceptor Pro+5(U) pICH41295 of Golden Gate MoClo Plant Tool Kit in a BbsI/T4 ligase reaction. This level 0 cloning is fully described in the master thesis of Magdalene Tsitsikle: "Mutagenesis of *Lotus japonicus* LSK-1 utilizing a CRISPR/Cas9 system".

## Level 0 coding sequences

Human codon-optimised Cas9 (**hCas9**) has been chosen for mutagenesis in *L. japonicus,* since it has been successfully used in other legume plants (Zhang, Yang, Yang, Li, & Guo, 2016). Because the original nucleotide sequence of hCas9 was quite large and, additionally, presented many internal recognition sites for BsaI and BbsI, six primer sets were designed to amplify it in six fragments from plasmid pK7WGF2::hCas9 (gifted to us by Dr. Athanasios Dalakouras). The domestication process for hCas9 coding sequence is explained below.

- Primers **Cas9_1F** and **Cas9_1R** amplified a fragment of 1423bp, including the start codon, while Cas9_1R introduced a synonymous mutation in position 3rd of BsaI recognition site 5'GAGACC3' (reverse complement of 5'GGTCTC3'), leading to 5'GAAACC3'. The coded amino acid at that position, glutamic acid (471) did not change.

- Primers **Cas9_2F** and **Cas9_2R** amplified a fragment of 290bp, while Cas9_2R introduced a synonymous mutation in position 3rd of BbsI recognition site 5'GAAGAC3', leading to 5'GAGGAC3'. The coded amino acid at that position, glutamic acid (566) did not change.

- Primers **Cas9_3F** and **Cas9_3R** amplified a fragment of 467bp, while Cas9_3R introduced a synonymous mutation in position 5th of BbsI recognition site 5'GTCTTC3', leading to 5' GTCTAC 3'. The coded amino acid at that position, leukine (720) did not change.

- Primers **Cas9_4F** and **Cas9_4R** amplified a fragment of 732bp, while Cas9_4R introduced a synonymous mutation in position 4th of BsaI recognition site 5'GGTCTC3', leading to 5' GGTGTC 3'. Coded valine (963) did not change.

- Primers **Cas9_5F** and **Cas9_5R** amplified a fragment of 869bp, while Cas9_5R introduced a synonymous mutation in position 5th of BsaI recognition site 5' GGTCTC 3', leading to 5' GGTCAC 3'. Coded serine (1248) did not change.

- Finally, primers **Cas9_6F** and **Cas9_6R** amplified a fragment of 392bp, including the stop codon.

Primers were designed to allow simultaneous assembly of fragments one, two and three in one level -1 acceptor and fragments four, five and six in another level -1 acceptor pAGM1311. Additionally, internal forward primers partially overlap with internal reverse primers to introduce mutations at the type IIS recognition sites and to generate compatible overhangs. Figure 3 presents with detail this design. Fragments were gel extracted and subcloned into level -1 acceptor. Cloning reaction was transformed into DH5a chemo-competent cells and transformed bacteria were plated onto kanamycin 50μg/ml LB medium plates. Blue-white

selection was applied and white colonies were chosen for plasmid isolation and diagnostic digestion. Double digestion was performed on a Cas9 frag 1-3 clone with PstI and SacI restriction enzymes. Expected zones were at 1323bp and 2879bp. A Cas9 frag 4-6 clone was double-digested with BstXI and XhoI, while expected zones were at 754bp and 3251bp (Fig. 3). Both clones were positive and were sent for Sanger sequencing to verify mutagenesis of internal recognition sites.

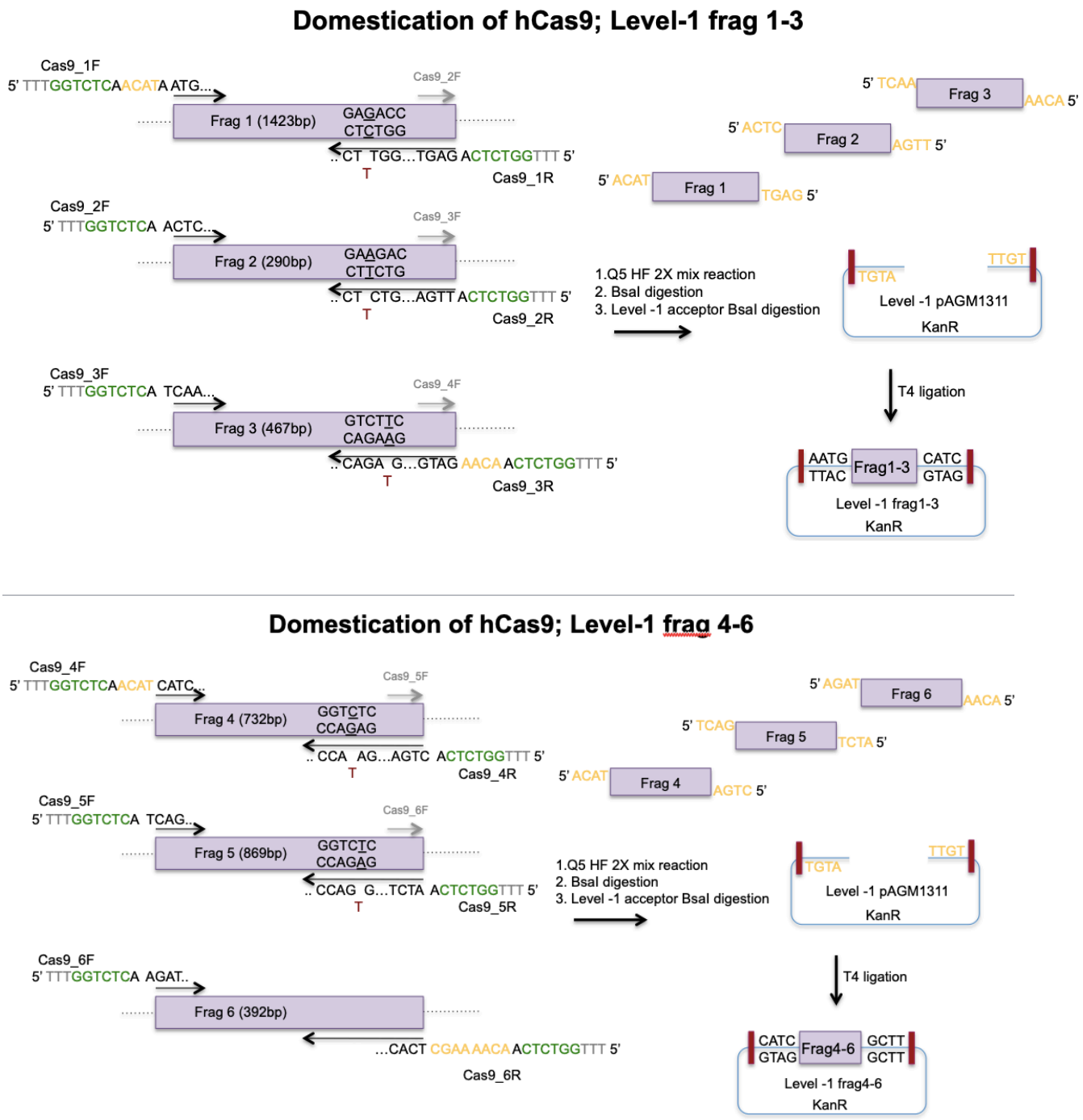



**Figure 3. Domestication and cloning of hCas9 coding sequence into two level -1 acceptor.**

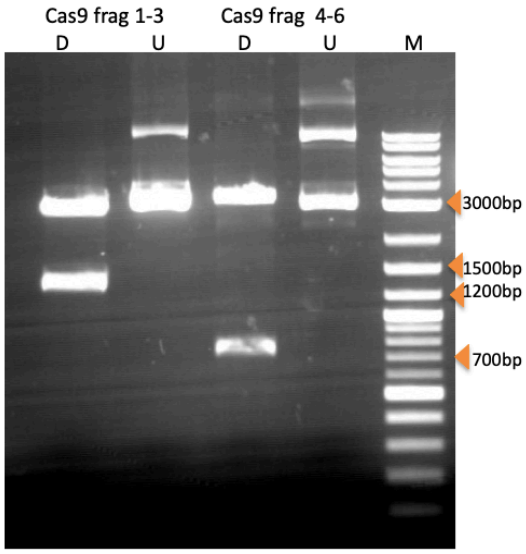



**Figure 13. Diagnostic digestion of level -1 Cas9 frag 1-3 and level -1 Cas9 frag 4-6 clones.** M: 2-Log DNA Ladder (0.1-10.0 kb), D: digested plasmid, U: uncut plasmid.

Subsequently, Cas9 fragments 1-3 and 4-6 were subcloned into level 0 acceptor CDS1 pICH41308 of Golden Gate MoClo Plant Tool Kit through a BbsI/T4 ligase reaction (Fig. 4). Cloning reaction was transformed into DH5a chemo-competent cells and transformed bacteria were plated onto spectinomycin 100μg/ml LB medium plates. Blue-white selection was applied and white colonies were chosen for plasmid isolation and diagnostic digestion. Double digestion was performed with SacI and XhoI, producing 1973bp and 4415bp (Fig. 15). Positive clone was sent for Sanger sequencing, whereas multiple consecutive chromatographs verified the expected nucleotide sequence.

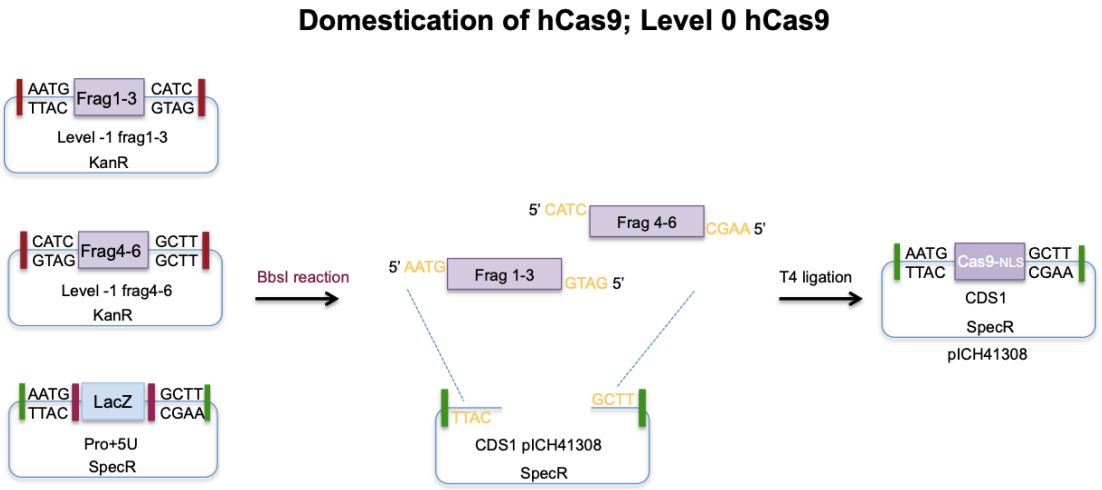



**Figure 4. Construction of level 0 Cas9 CDS1 module**.

Apart from level 0 Cas9 module, other coding sequences are required for efficient selection of transgenic tissues. We decided to employ two selection schemes on our final CRISPR/Cas9 binary vector, a reporter cassette and a selection marker on each side of the T-DNA region. We selected vector pICSL80005 from Golden Gate MoClo Plant Parts Kit, a level 0 CDS1 module, harbouring a variant of **turboGFP (**from *Pontellina plumata*), codon-optimized for expression in plants. Regarding the selection cassette, we decided to utilize the **Bar** gene encoding for phosphinothricin acetyltransferase. This enzyme makes transgenic plants

resistant to herbicide PPT. Bar coding sequence was amplified from plasmid pICSL70005, using primers Bar_F and Bar_R in a Q5® High-Fidelity 2X mix reaction. The 5'overhangs of these primers introduced a BbsI recognition site, followed by the appropriate 4-bp sequence, 5'AATG3' on the left side and 5'GCTT3' on the right side (Fig. 6). After PCR clean up, Bar amplicon was cloned into level 0 acceptor CDS1 pICH41308 in a BbsI/T4 ligase reaction (Fig. 6). Cloning reaction was transformed into DH5a chemo-competent cells and transformed bacteria were plated onto spectinomycin 100μg/ml LB medium plates. Blue-white selection was applied and a level 0 Bar clone was verified through diagnostic digestion and Sanger sequencing. Level 0 Bar clone was digested with StuI, giving a single 2800bp band (Fig. 5).

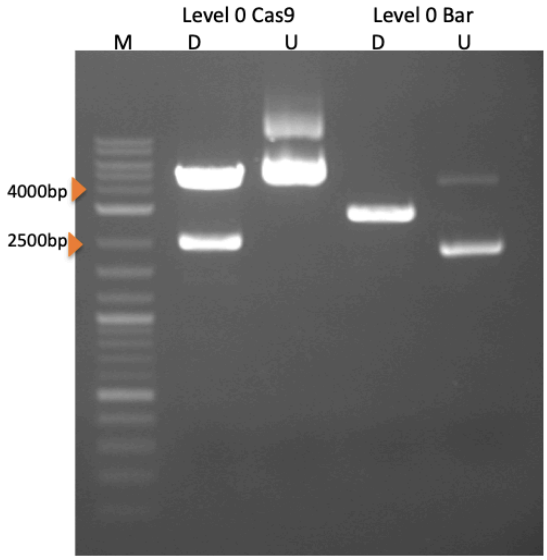



**Figure 5. Diagnostic digestion of level 0 Cas9 and level 0 Bar clones**. M: 2-Log DNA Ladder (0.1-10.0 kb), D: digested plasmid, U: uncut plasmid.

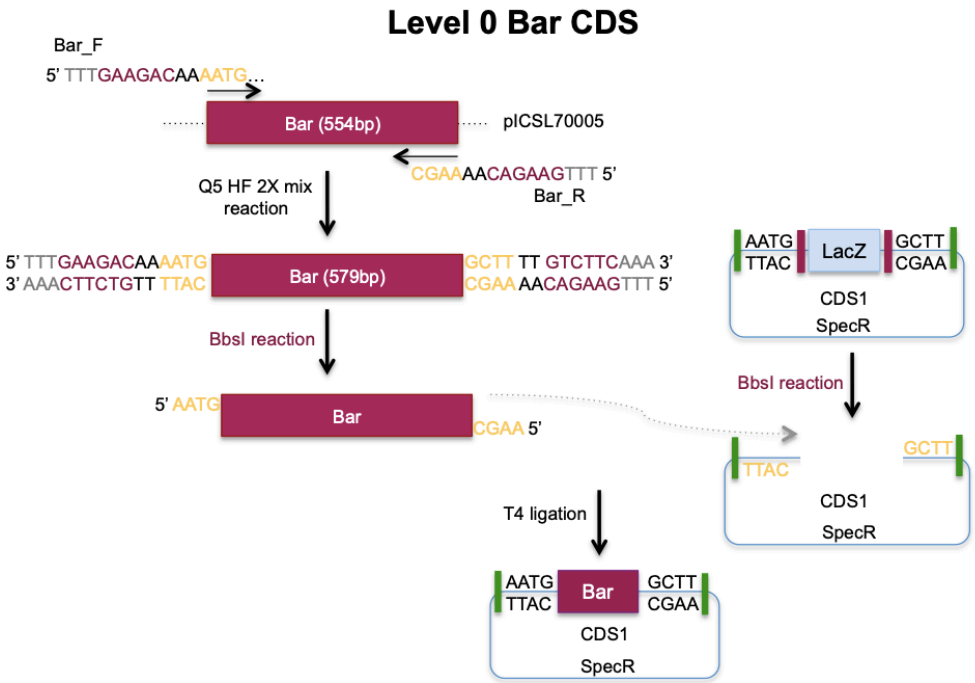



**Figure 6. Construction of level 0 Bar CDS1 module.**

## Level 0 terminators

For our CRISPR/Cas9 level 0 collection, we utilized two level 0 3U+Ter modules, pICH41414 and pICH44300 from Golden Gate MoClo Plant Parts Kit. The first vector harbours a 3'UTR and polyadenylation signal/terminator of **CaMV 35s** gene, while the second vector consists of a 3'UTR and polyadenylation signal/terminator of an actin gene (**Act2**) from A. *thaliana*.

## Level 0 amino terminal tags/dsDNA oligo

To ensure efficient translocation of Cas9 protein to the plant nucleus, an amino terminal nuclear localization signal was needed. N' nuclear localization signal of Simian Virus 40 (**N'NLS SV40**) was synthesised in the form of two single stranded oligonucleotides, N'SV40NLS_F and N'SV40NLS_R, that were partially complimentary to each other. After oligonucleotide annealing, N' NLS dsoligo was flanked by the same sites as with a level 0 NT1 module, meaning 5'CCAT 3' on the left side and 5'AATG 3' on the right side (Fig. 7). We decided to utilize it as dsoligo, instead of level 0 module to increase cloning efficiency. Nevertheless, it was subsequently subcloned into level 0 NT1 acceptor pAGM1276 for storing.

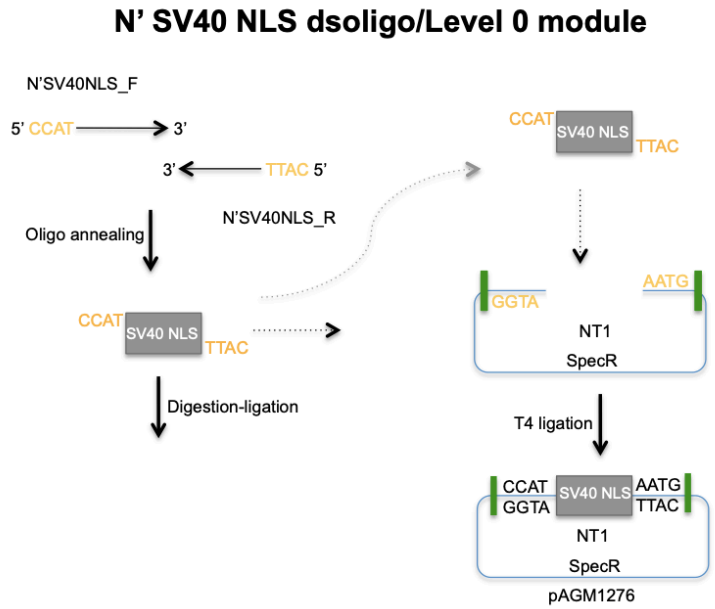



**Figure 7. Dsoligo N' SV40 NLS formation. Subsequent cloning into level 0 NT1 acceptor.**

## GuideRNA scaffold level 0 module

The sequence of guideRNA (gRNA) scaffold was amplified from vector-pich86966::AtU6p::sgRNA_DCR2 and subcloned into level 0 universal vector pAGM9121. The gRNA scaffold can then be released by BsaI digestion and the fragment will be flanked by 5'AATG3' on the left side and 5'CGCT3' on the right side. The left overhang is compatible with the right overhang of level 0 LjU6 promoter and is amenable to assembly into level 1 acceptors. This level 0 cloning is fully described in the master thesis of Magdalene Tsitsikle: "Mutagenesis of *Lotus japonicus* LSK-1 utilizing a CRISPR/Cas9 system".

All the aforementioned level 0 modules comprise the CRISPR/Cas9 level 0 collection. Figure 8 summarizes its contents.

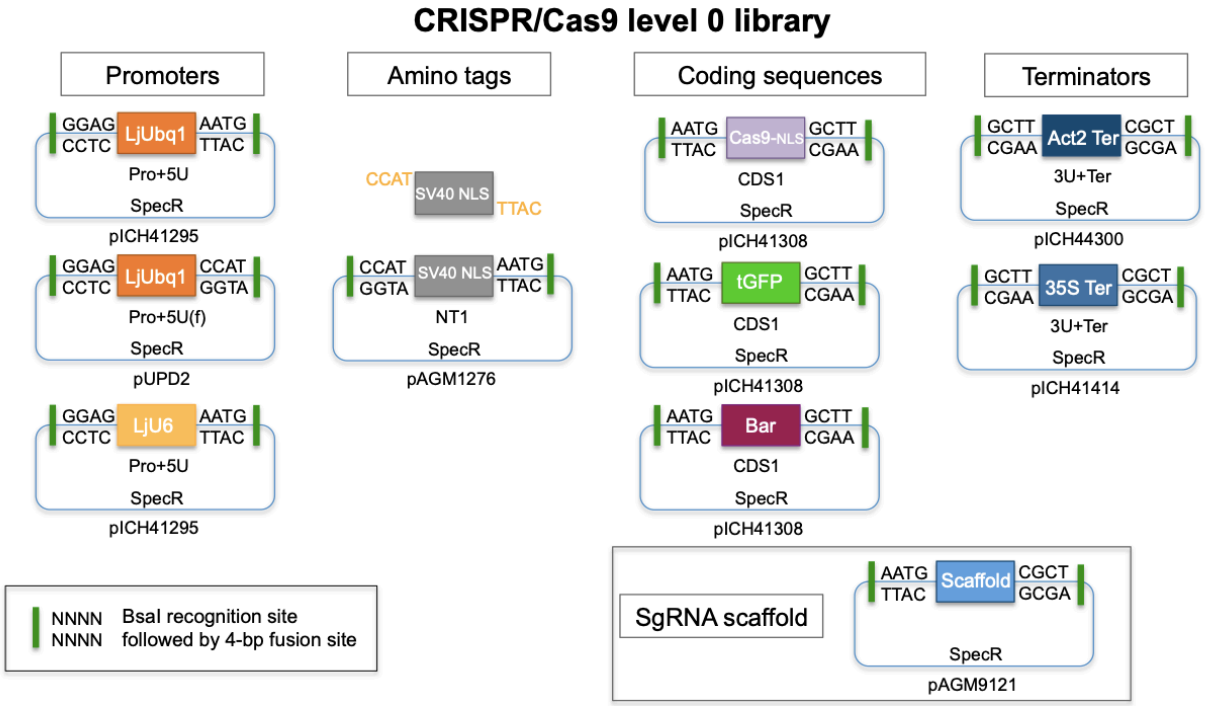



**Figure 88. CRISPR/Cas9 level 0 collection.**

## 1.2. CRISPR/Cas9 level 1 collection of transcriptional units

In Golden Gate cloning, the order and orientation of transcriptional units in downstream high-level multigene constructs should be considered before assembling level 0 modules into level 1 acceptor backbones. To optimise CRISPR/Cas9 gene editing efficiency, we considered a list of different factors, like Cas9 sequence, promoters to drive the expression of Cas9 and sgRNA, combination between promoter and terminator and finally T-DNA architecture. How transcriptional units are organized within the T-DNA borders seems to play a role in enhancing mutation efficiency and even allowing for credible visual recognition of transgenic tissues. As Golden Gate cloning allows different combinations regarding the order (1st, 2nd, 3rd position, etc) and orientation (forward or reverse) of the cassettes, we selected the one for which evidence exist supporting its enhanced gene editing effect.

Cas9 expression cassette and gRNA expression cassette(s) are located in the middle of the T-DNA region in opposite directions. Cas9 cassette is located in the 2nd position to facilitate a simpler cloning scheme. GuideRNA cassette(s) follow behind at the 3rd and optionally 4th position. In the language of Golden Gate cloning, a transcriptional unit follows the forward orientation, if its promoter is closer to the left border, and the reverse orientation, if it is closer to the right border (Weber et al., 2011). With that in mind, Cas9 cassette was subcloned into level 1 acceptor of the 2nd position and reverse orientation. GuideRNA acceptors, which are able to accept a gRNA sequence, were constructed for the 3rd and 4th position in forward orientation. This ‘head-to-head’ orientation seemed to enhance mutation efficiency in Arabidopsis (Castel, Tomlinson, Locci, Yang, & Jones, 2019). A simple explanation is that a weak terminator after Cas9 enables RNA-polymerase II readthrough within the gRNA expression cassette, preventing optimal expression of the gRNA. Thus, using a strong terminator (e.g. CaMV 35s) and expressing Cas9 and gRNA in opposite directions is recommended.

A reporter or selection cassette can delimit each T-DNA border. We decided to enlist a reporter cassette close to the left border and a plant selection marker at the right border. Since incorporation of T-DNA region into the plant genome has polarity, from right to left border, (Bahramnejad, Naji, Bose, & Jha, 2019) an easily identified reporter gene is ideal for that position. A selection marker is also enlisted, because having two selection schemes increases chances of identifying true transgenic tissues, especially in stable transformation. Keeping these principles in mind, we assembled the following level 1 transcriptional units.

For the construction of the reporter cassette, level 0 modules LjUbq1 promoter (Pro+5U), turboGFP coding sequence and terminator of actin 2 gene were subloned into level 1 acceptor pICH41337 (1st position and

forward orientation, referred to as 1F) through a BsaI/T4 ligase reaction (Fig. 9). Cloning reaction was transformed into DH5a chemo-competent cells and transformed bacteria were plated onto carbenicillin 100μg/ml LB medium plates. Blue-white selection was applied and a white colony was chosen for plasmid isolation. The clone was simultaneously digested with SpeI and BsaXI, while gel electrophoresis revealed the expected digestion band motif, two visible bands at 856bp and 5817bp, respectively (Fig. 10). A band of 30bp is non-visible. The positive clone was also verified through Sanger sequencing.

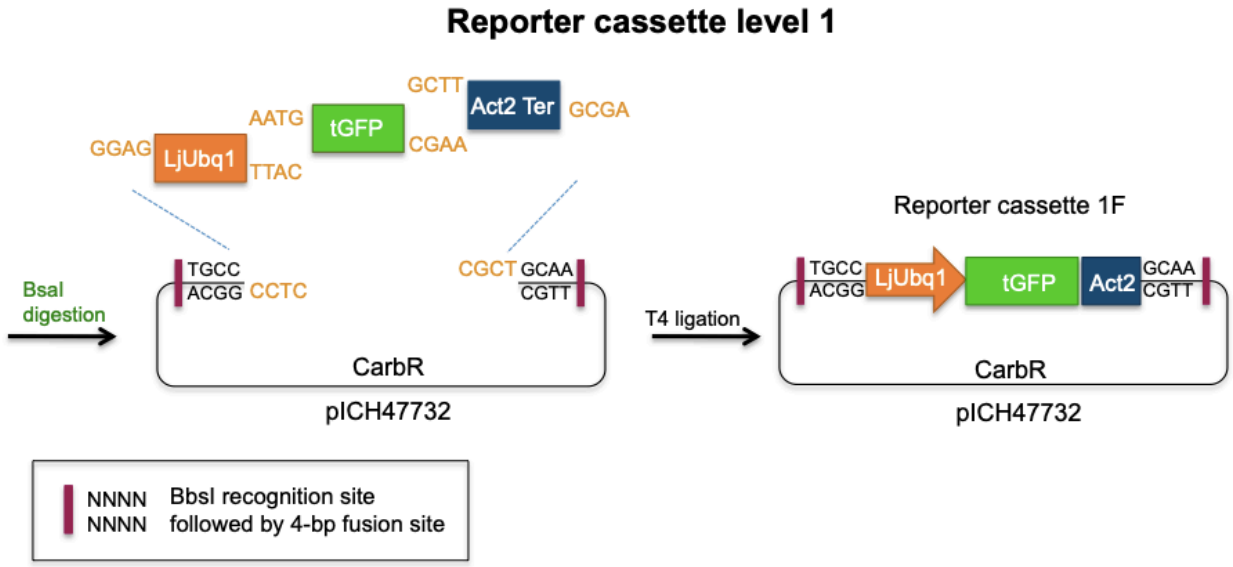



**Figure 9. Assembly of level 1 reporter cassette**.

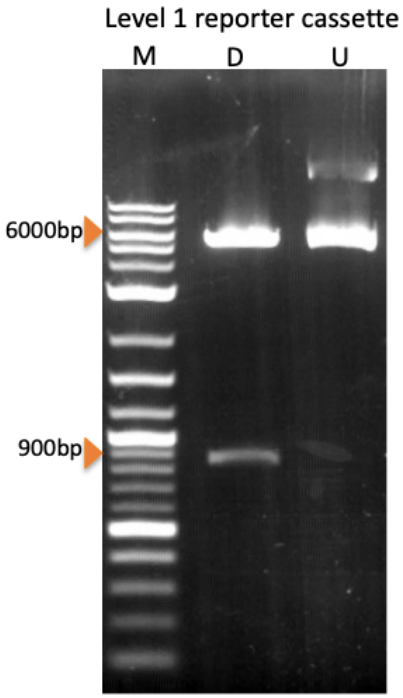



**Figure 10. Diagnostic digestion of level 1 reporter cassette.** M: 2-Log DNA Ladder (0.1-10.0 kb), D: digested plasmid, U: uncut plasmid.

The **Cas9 expression cassette** was constructed in a similar manner. Level 0 modules LjUbq1 (Pro+5U(f)), ds oligo N' SV40 NLS, Cas9 coding sequence and CaMV 35S terminator were subcloned into level 1 acceptor pICH47811 (2nd position and reverse orientation, **2R**) (Fig. 11). After transformation and plasmid

isolation, double diagnostic digestion was performed with EcoRI and XmaI. The expected digestion band motif (536bp, 1664bp, 2456bp and 5239bp) can be seen in figure 12.

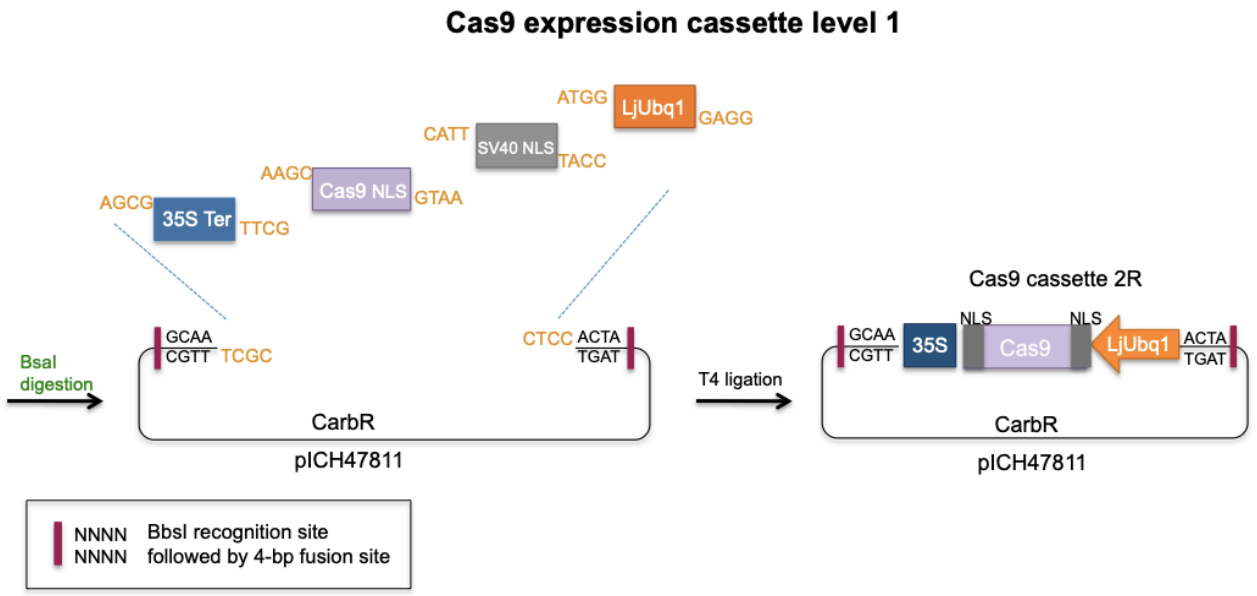



**Figure 11. Assembly of level 1 Cas9 expression cassette.**

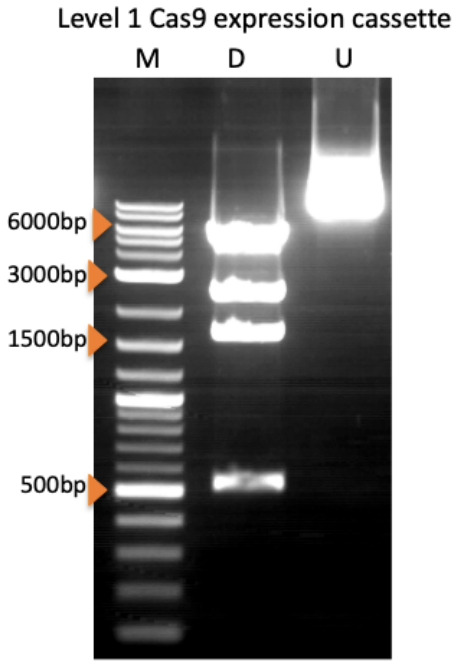



**Figure 12. Diagnostic digestion of level 1 Cas9 expression cassette**. M: 2-Log DNA Ladder (0.1-10.0 kb), D: digested plasmid, U: uncut plasmid.

Our laboratory has decided on the following scheme for the incorporation of the gRNA cassette(s). To create a series of **gRNA acceptors**, vectors that can accept gRNA sequences, for different positions downstream of the Cas9 expression cassette. So far, we have designed and utilized gRNA acceptors for the 3$^{rd}$ and 4$^{th}$ position. Level 0 modules LjU6 promoter and gRNA scaffold were subcloned into level 1 acceptors pICH47751 (**3F**) and pICH47761 (**4F**) (Fig. 13). After transformation and plating onto carbenicillin 100μg/ml plates, white colonies were selected for plasmid isolation. A single colony was chosen for each construct and both of them were digested with PsiI. Two bands at 1537bp and 3776bp

verified the successful cloning reactions (Fig. 14). Both clones were sent for Sanger sequencing, further confirming their nucleotide sequence. In the future and, depending on our needs, we can easily assemble gRNA acceptors for additional positions.

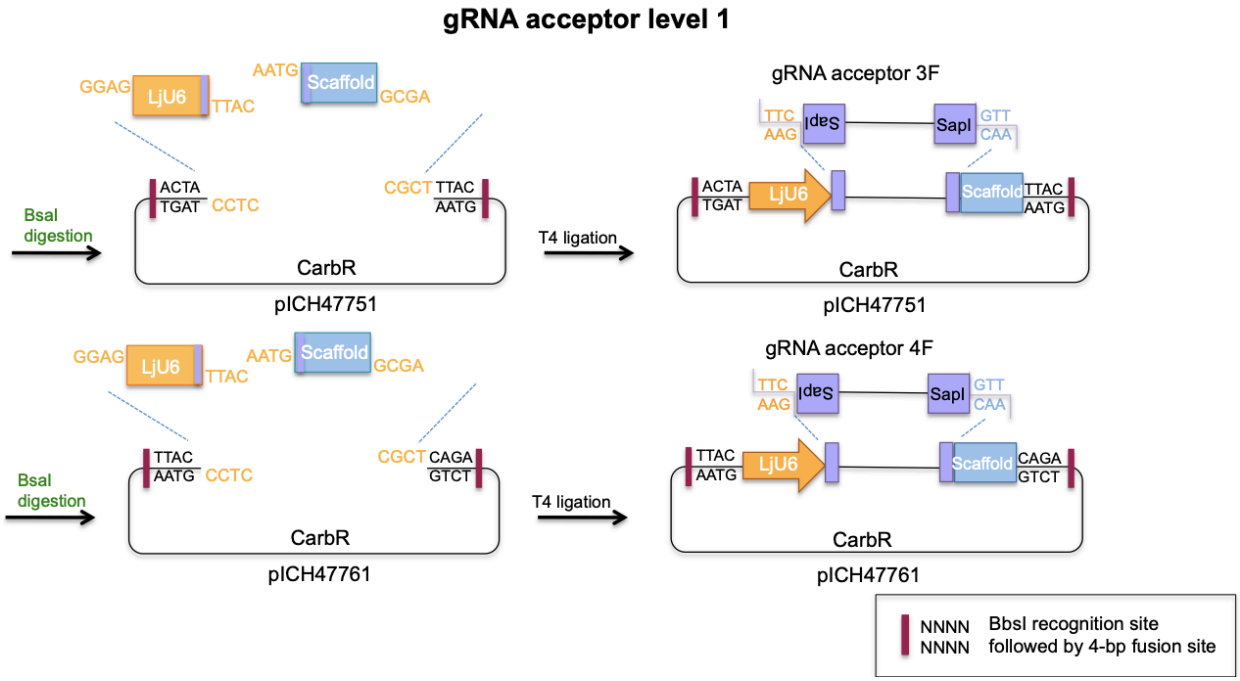



**Figure 13. Assembly of level 1 gRNA acceptor.**

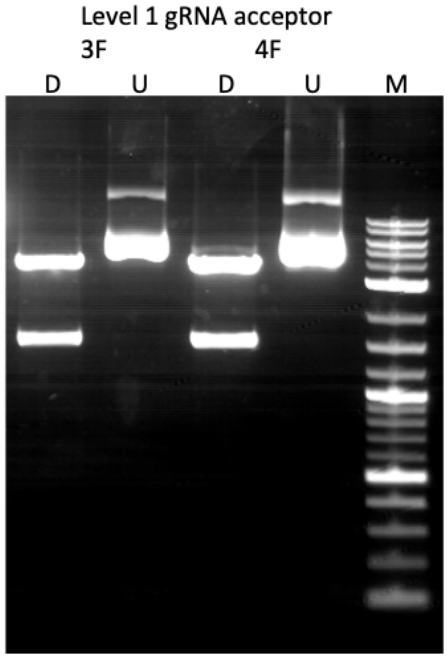



**Figure 14. Diagnostic digestion of level 1 gRNA acceptor 3F and 4F**. M: 2-Log DNA Ladder (0.1-10.0 kb), D: digested plasmid, U: uncut plasmid.

Finally, based on the number of gRNA cassettes in the final CRISPR/Cas9 binary vector, a plant selection marker of the complimentary position has to be used. Currently, we have the **selection marker cassette** available for the 4$^{th}$ and 5$^{th}$ position in reverse orientation (**4R** and **5R** respectively). Level 0 modules LjUbq1 promoter (Pro+5U), Bar coding sequence and CaMV 35S terminator were subcloned into level one acceptors

pICH47831 (**4R**) and pICH47841 (**5R**) (Fig. 15). After transformation and plating onto carbenicillin 100μg/ml plates, white colonies were selected for plasmid isolation. Double diagnostic digestion was performed in both constructs with BsaAI and XhoI producing three bands at 309bp, 947bp and 5004bp (Fig. 16). We also confirmed each construct through Sanger sequencing. In the future, if more gRNA cassettes are to be used in the final vector, we will also construct a level 1 selection marker of the appropriate position in proximity to the right border.

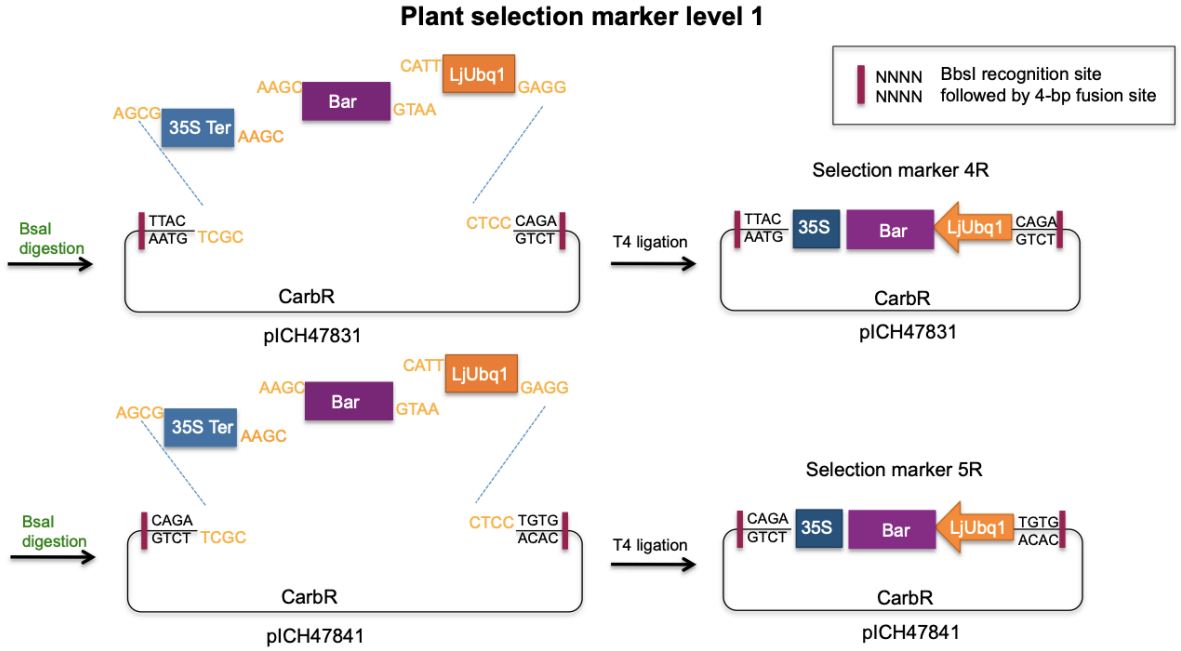



**Figure 15. Assembly of level 1 selection marker.**

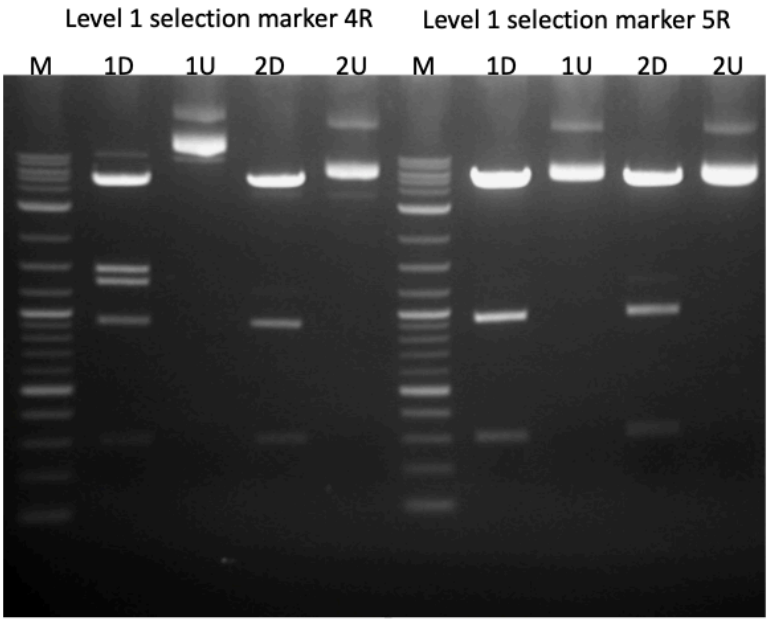



**Figure 16. Diagnostic digestion of level 1 selection marker 4R and 5R.** M: 2-Log DNA Ladder (0.1-10.0 kb), D: digested plasmid, U: uncut plasmid. 1, 2: number of white colonies that were analysed. Level 1 selection marker 4R colony 1 was wrong. Level 1 selection marker 4R colony 2 and level 1 selection marker 5R colony 1 were sent for Sanger sequencing.

The CRISP/Cas9 level 1 collection of transcriptional units can be employed for gene editing of any genomic locus in *L. japonicus*. During the duration of this master thesis, three genes were targeted for CRISPR/Cas9-

mediated gene editing, AMY2, CYP71D353 and LSK1.

## 2. CRISPR/Cas9 knockout of *L. japonicus* AMY2, CYP71D353 and LSK1 genes

We aim to employ the described CRISPR/Cas9 system to inactivate two different types of genes, *AMY2* and *CYP71D353* encoding for triterpene synthases and *LSK1* encoding for a kinase, for future functional analyses. Gene inactivation, or more specifically reduced protein function, requires the open reading frame to be disrupted. This can also be caused by a premature or later stop codon that will produce an unnaturally short or long protein chain. We can certainly achieve gene inactivation by targeting a Cas9/gRNA complex to the coding region, meaning the exons, preferably the ones closer to the 5'UTR. We can also target coding regions that encode for protein active domains (Shan, Soltis, Soltis, & Yang, 2020). However, this method has some limitations regarding mutagenesis of a gene with high homology to others. In this case, two gRNAs can be used simultaneously to target the less homologous introns, thus removing the intermediate exonic region. Notably, the removal of a single or multiple exons can be more effective in decreasing protein functionality, since a large number of amino acid residues will be absent from the produced protein. Following the same logic, it is best to employ multiple gRNAs targeting different exons. This approach is beneficial for two reasons. First, the CRISPR/Cas9 system can introduce either indels at each exon target site or/and lead to deleted regions. Second, the chances of successful gene editing are higher, since not all designed gRNAs will be equally efficient in generating mutations. In fact, many of them may be unable to introduce any mutation at all. Taking all these into consideration, we selected the following target sites for CRISPR/Cas9 gene editing.

### 2.1. Selection of target sites

Genomic sequences and coding sequences for *AMY2*, *CYP71D353* and *LSK1* were acquired from Lotus Base or/and Genbank. Two different CRISPR/Cas9 design tools were used, CRISPR-P and CRISPRdirect, since each tool takes into consideration different parameters to score each gRNA sequence. For ***AMY2***, **five target sites** were selected with three of them located in introns and the other two in the same exon (Fig. 17). We selected the three gRNAs AMY2-gRNA1, AMY2-gRNA4 and AMY2-gRNA5 to target intron one, nine and ten respectively, because AMY2 is highly homologous to AMY1, which is another triterpene biosynthetic gene. AMY2-gRNA2 and AMY2-gRNA3, which target exon four, were one of few with a low number of off-target sites and relatively high on-score, respectively. ***CYP71D353*** was targeted by **two gRNAs**, CYP71D353-gRNA1 and CYP71D353-gRNA2. Since this triterpene biosynthetic gene consists of two large exonic regions, both of them gRNAs target the first exon. Last, for ***LSK1***, **four target sites** were selected, all in different exons. LSK1-gRNA1 targets exon three, LSK1-gRNA2 targets exon four, while LSK1-gRNA3 and LSK1-gRNA4 target exon five and ten, respectively.

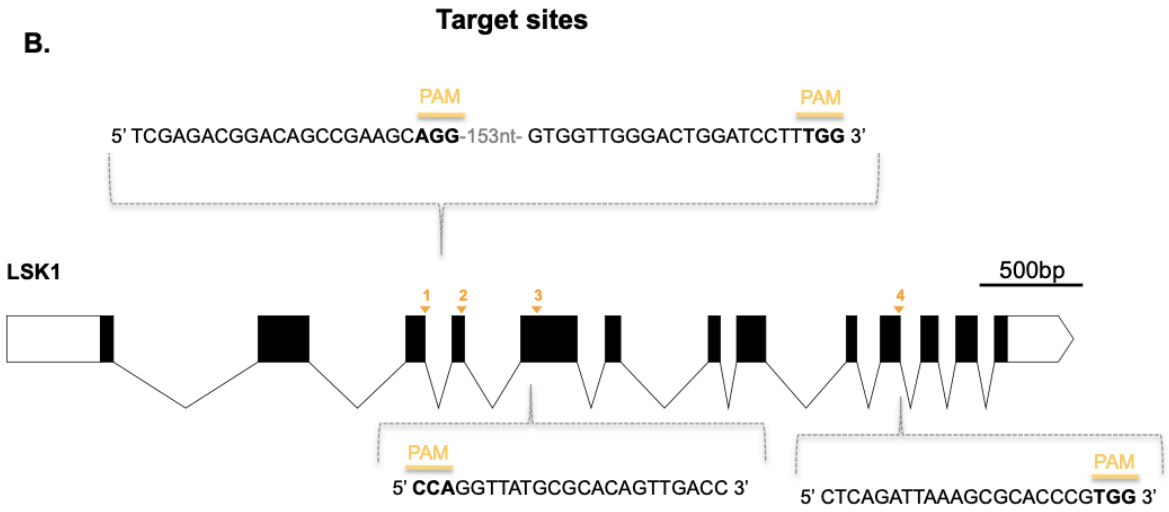



**Figure 17. CRISPR/Cas9 target sites in AMY2, CYP71D353 and LSK1 genes.**

We decided to test the mutation efficacy of each gRNA via *Agrobacterium rhizogenes*-mediated hairy root transformation. This approach can be used for nodulation and symbiotic nitrogen fixation related studies, but at this point we were interested in assessing the mutation efficiency of our CRISPR/Cas9 system. We decided against using each gRNA separately, but combined them as to target each time a different gene. The number and combination of gRNAs utilized in each hairy root transformation was determined after careful consideration. First, the majority of the CRISPR/Cas9 binary vectors expressed two gRNAs targeting the same gene, thus allowing fewer numbers of hairy root transformations and easier identificaton of gene editing if both gRNAs were to be functional. Additionally, a intron targeting-gRNA should be at least combined with another, in order to disrupt the open reading frame. Secondly, we used a maximum of two gRNA cassettes, because analysis in other plant species has shown that mutation efficacy decreases as the number of expressed gRNAs increases. This was especially observed when Cas9 expression was driven under a strong constitutive promoter. Lastly, the distance between each target site can also affect the mutation efficiency, since two Cas9/gRNA complexes may "compete" for the same targeted region (Hashimoto et al., 2018). For these reasons, as presented in the next section "2.2 Construction of gRNA expression cassettes", AMY2-gRNA2 was used in combination with AMY-gRNA1 and on its own. AMY2-gRNA3 was also used by itself, in case that AMY2gRNA1 is unffective in introducing indels. AMY-gRNA4 and AMY-gRNA5 created the fourth vector. CYP71D353-gRNA1 and CYP71D353-gRNA2 were combined in the same vector. For LSK1, two vectors were constructed, one harboring LSK1-gRNA1 and LSK1-gRNA3 and the other LSK1-gRN2 and LSK1-gRNA4.

## 2.2. Construction of gRNA expression cassettes

Each gRNA will be produced by a gRNA expression cassette. gRNA acceptors 3F and 4F harbor two SapI recognition sites between the LjU6 promoter and gRNA scaffold. After digestion with SapI, two overhangs are produced, 3'AAC5' and 5'GTT3'. Consequently, gRNAs are synthesised as oligonucleotides with 5' overhangs complimentary to those of the gRNA acceptor. After annealing of oligonucleotides, T4 ligation produces each gRNA cassette (Fig. 18). Based on the selected combinations of expressed gRNAs in the final CRISPR/Cas9 vectors, each dsoligo gRNA was subcloned into a specific gRNA acceptor.

- AMY2-gRNA1_ F/R in gRNA acceptor 4F (4$^{rd}$ posititon and forward orientation)
- AMY2-gRNA2_ F/R in gRNA acceptor 3F (3$^{rd}$ posititon and forward orientation)
- AMY2-gRNA3_ F/R in gRNA acceptor 3F
- AMY2-gRNA4_ F/R in gRNA acceptor 3F
- AMY2-gRNA5_ F/R in gRNA acceptor 4F
- CYP71D353-gRNA1_ F/R in gRNA acceptor 3F
- CYP71D353-gRNA2_ F/R in gRNA acceptor 4F

- LSK1-gRNA1_ F/R in gRNA acceptor 3F
- LSK1-gRNA2_ F/R in gRNA acceptor 3F
- LSK1-gRNA3_ F/R in gRNA acceptor 4F
- LSK1-gRNA4_ F/R in gRNA acceptor 4F

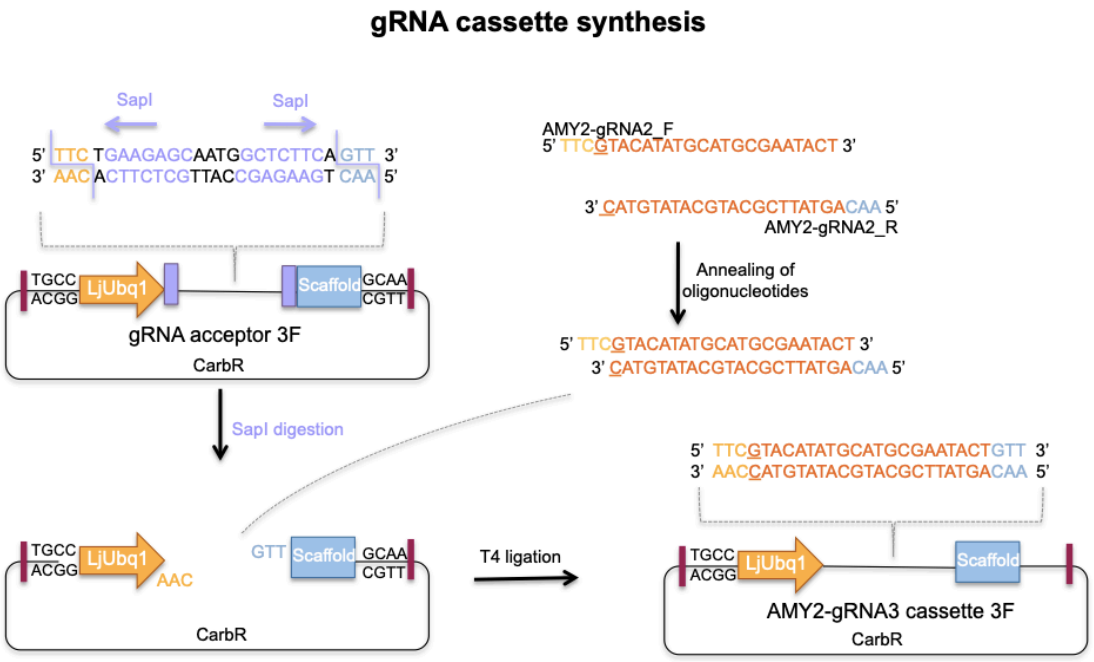



**Figure 18. Assembly of gRNA cassettes**. Cloning of dsoligo AMY2-gRNA3 in gRNA acceptor 3F is depicted as an example. All gRNA cassettes were made following this workflow.

After transformation, tranformed bacteria were plated onto carbenicillin 100μg/ml plates. Since gRNA acceptors do not have a lacZ cassette, blue-white selection was not possible. Nevertheless, the digestion-ligation protocol is very efficient and one to two colonies are enough to give a positive clone. Diagnostic PCR was used to verify the insertion of some gRNAs in the vector, using the reverse primer of each dsoligo and a forward primer that binds to LjU6p. The exact nucleotide sequence of each gRNA in the gRNA acceptors was confirmed for all gRNA cassettes via Sanger sequencing

## 2.3. Assembly of complete CRISPR/Cas9 binary vectors

Level 1 transcriptional unit reporter cassette, Cas9 expression cassette, gRNA cassette(s) and plant selection marker were assembled into level 2 vector pAGM4723 in a BbsI/T4 ligase reaction (Fig. 19). The appropriate end-linker was used to connect the far right selection marker to the right overhang of pAGM4723. A vector containing everything but the 20nt sequence(s) of gRNA(s) was also assembled as the negative control. It was constructed by replacing the gRNA cassettes with the gRNA acceptors and maintaining everything else. Because the assembly of the CRISPR/Cas9 binary vectors requires simultanous digestion of seven or six parts and immediate ligation of the inserts into the backbone, we introduced some modifications to our cloning and transformation protocols (see sections: Materials and Methods 1. Heat shock transformation protocol with DH5a chemo-competent bacterial cells and 2. Golden Gate-based plasmid construction). All the volume (10μl) of each cloning reaction was transformed into DH5a chemo-competent cells and transformed bacteria were plated onto kanamycin 25μg/ml plates. Because vector pAGM4723 carries a canthaxanthin biosynthesis operon (Cred), red-white selection can be applied. White colonies are expected to harbor all transcriptional units with the appropriate order. However, T4 ligase can sometimes ligate non-complimentary overhangs, producing vectors without the full number of inserts. This is why for each construct we screened two-five white colonies instead of the usual one-two. All CRISPR/Cas9 vectors were digested with XmnI. There are four XmnI recognition sites in each vector, one in the turboGFP coding sequence, one in the Cas9 coding

sequence and one in the LjU6 promoter. Therefore, all vectors harboring two gRNA cassettes and the negative control are expected to produce four bands: 2189bp, 4345bp, 962bp and 8924bp (Fig. 20). The two CRISPR/Cas9-AMY2 vectors carrying one gRNA cassette should give three bands: 2189bp, 4345bp and 8924bp (Fig. 21). All constructs were verified through diagnostic digestion and Sanger sequencing.

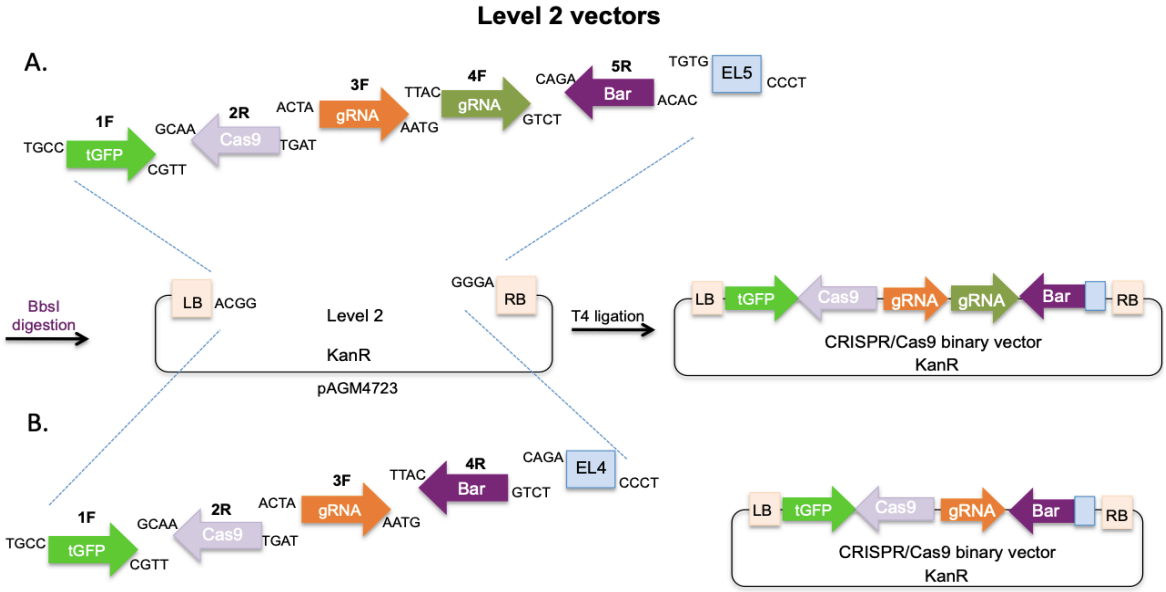



**Figure 19**. **Assembly of level 2 CRISPR/Cas9 binary vectors**.

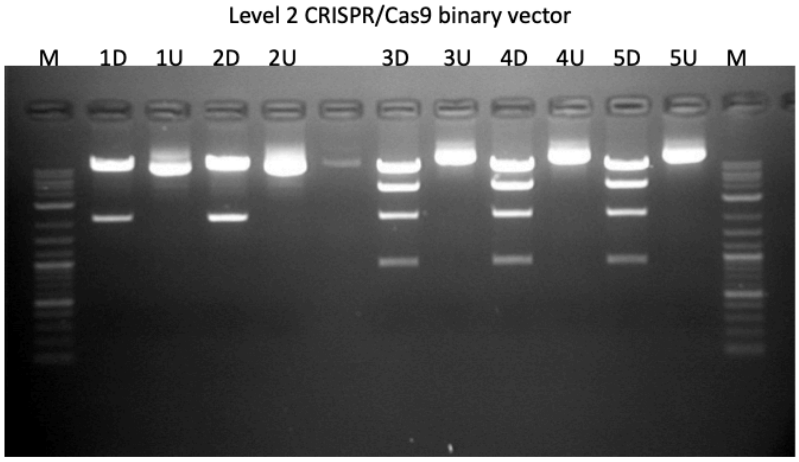



**Figure 20. Diagnostic digestion of level 2 vector.** AMY:gRNA4-gRNA5 is depicted as an example. M: 2-Log DNA Ladder (0.1-10.0 kb), D: digested plasmid, U: uncut plasmid. 1-5: number of white colonies that were analysed. Clones 3-5 were correct.

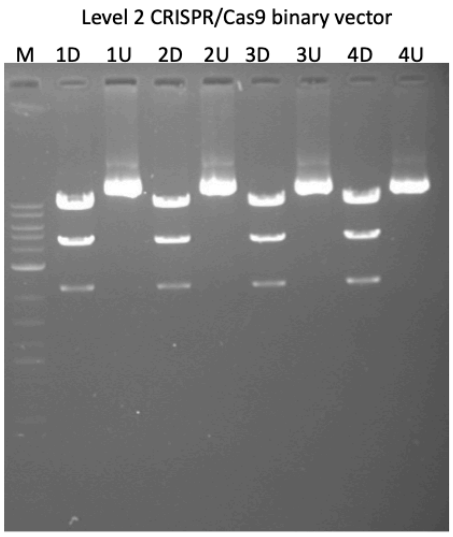



**Figure 21. Diagnostic digestion of level 2 vector.** AMY:gRNA3 is depicted as an example. M: 2-Log DNA Ladder (0.1-10.0 kb), D: digested plasmid, U: uncut plasmid. 1-4: number of white colonies that were analysed. All four clones were correct.

## 2.4. *Agrobacterium rhizogenes*-mediated hairy root transformation

CRISPR/Cas9 binary vectors were introduced into *A. rhizogenes* LBA 1334 strain by electroporetion. Even though eight CRISPR/Cas9 binary vectors were generated, including one level 2 Cas9-only, two were not utilized in the hairy root transformations of this master thesis. These vectors, AMY2:gRNA2, AMY2:gRNA3, will be employed in future experiments. Utilized CRISPR/Cas9 binary vectors and targeted gene regions are presented below (Table 1).

**Table 1. List of utilized CRISPR/Cas9 binary vectors**

| Target gene region | CRISPR/Cas9 binary vectors | Vector size | T-DNA size |
|---|---|---|---|
| AMY2: intron 1 and exon 4 | AMY2:gRNA1-gRNA2 | 16420bp | 11894bp |
| AMY2: intron 9 and intron 10 | AMY2:gRNA4-gRNA5 | 16420bp | 11894bp |
| CYP71D353: exon 1 | CYP71D353:gRNA1-gRNA2 | 16420bp | 11894bp |
| LSK1: exon 3 and exon 5 | LSK1:gRNA1-gRNA3 | 16419bp | 11893bp |
| LSK1: exon 4 and exon 10 | LSK1:gRNA2-gRNA4 | 16419bp | 11893bp |
| - | Cas9-only (negative control) | 16418bp | 11892bp |

By comparing the phenotypic characteristics of hairy roots (size, length, number of vertical roots) between control and CRISPR/Cas9-exposed lines, we do not observe any notable difference (Fig. 22). At this stage, we did not know if these specific roots were transformed with the T-DNA region. However, during root sampling the majority, if not all, of hairy roots were GFP positive. A very obvious explanation is that *AMY2*, *CYP71D353* and *LSK1* are mainly expressed during inoculation with rhizobium. Therefore, a potential gene-editing event would not be phenotypically observed, since AMY2, CYP71D353 and LSK1 protein expression remains at low levels during normal development. Another possibility is that none of the CRISPR/Cas9 systems introduced any mutation. The only way to be sure is to search for gene editing effects in the targeted gene regions. Hairy roots exhibiting GFP fluorescence were sampled (Fig. 23).

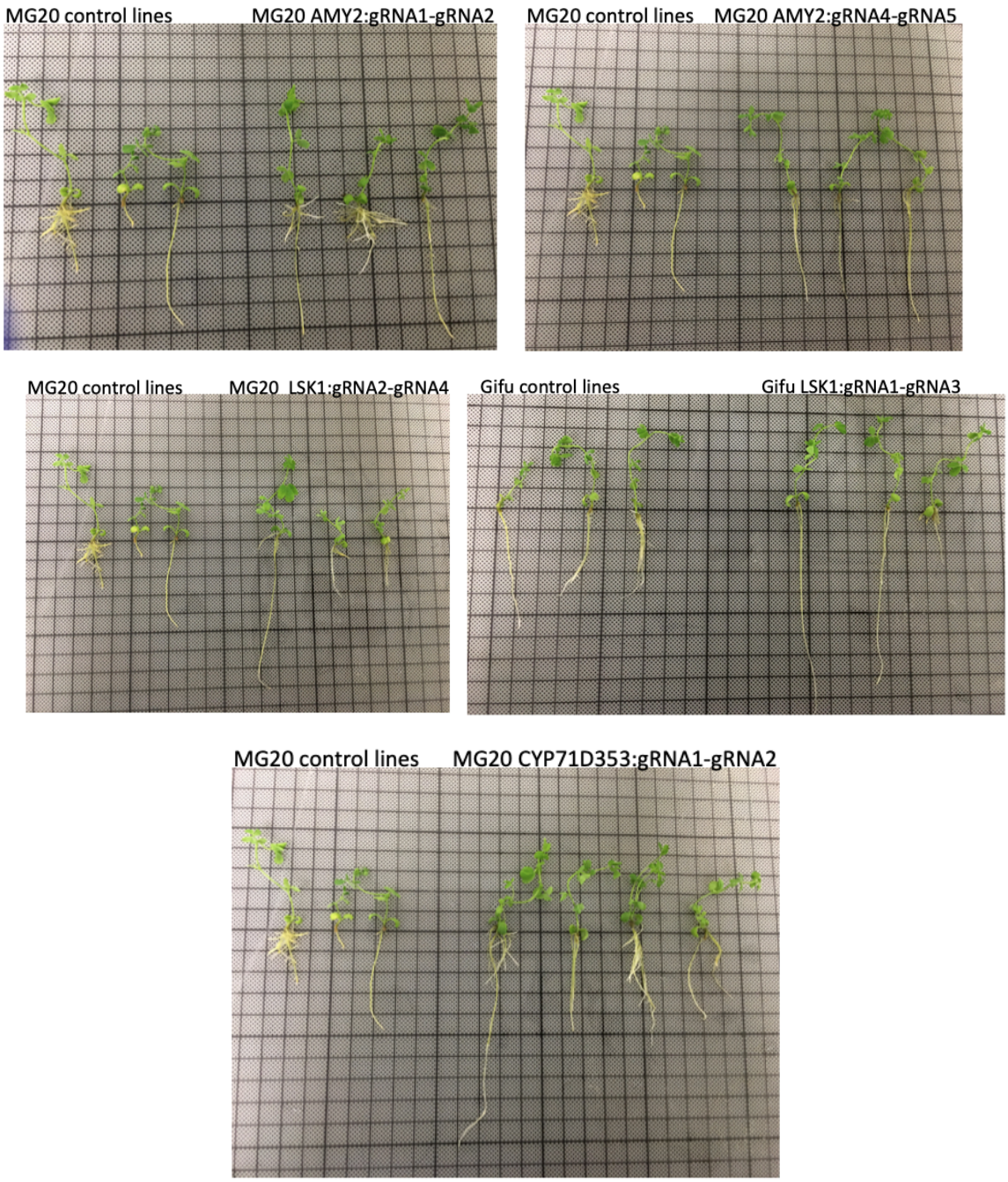



**Figure 22. Phenotypic comparison between control and CRISPR/Cas9-exposed lines.**

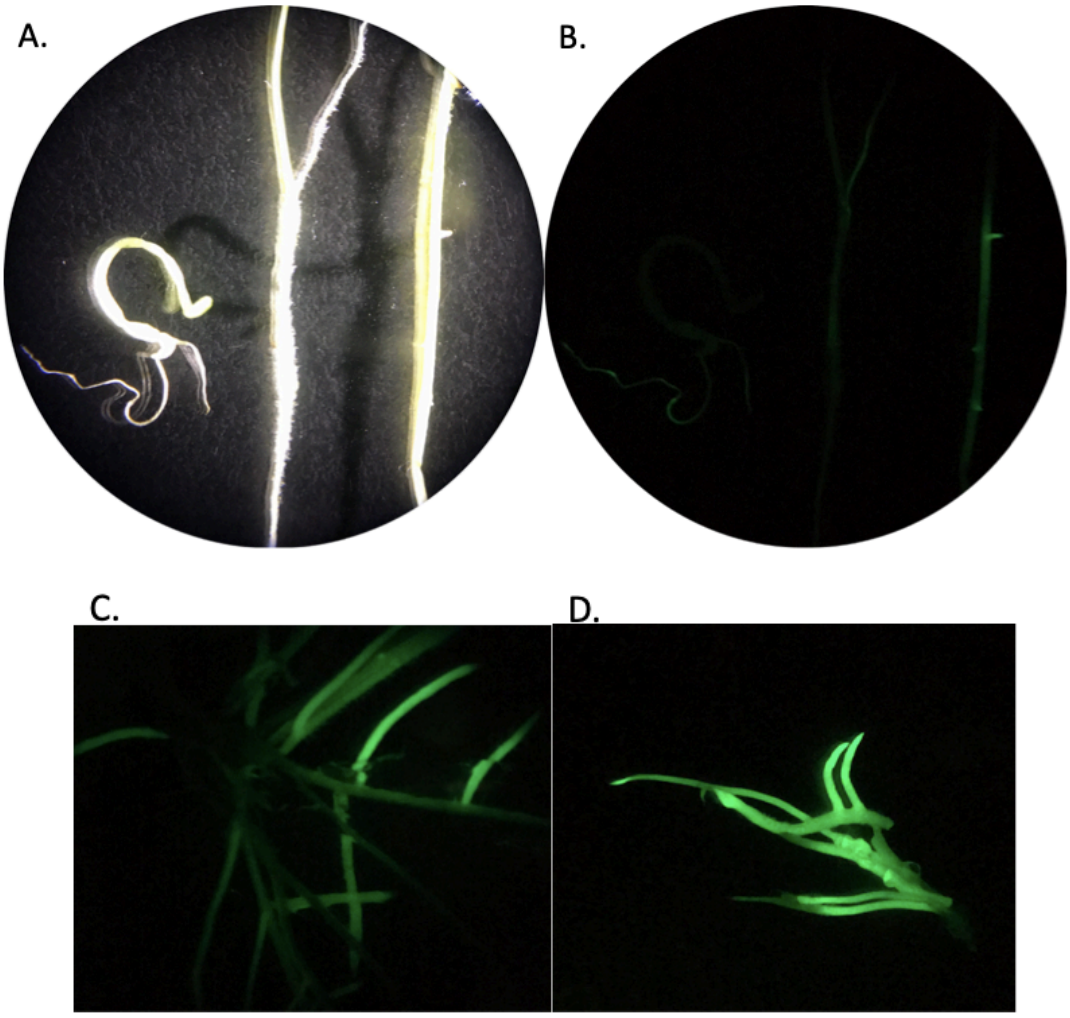



**Figure 23. GFP positive hairy roots under stereoscope.** A) and B) show three roots, left is wild type, middle and right are transformed with CRISPR/Cas9 T-DNA region. C) and D) show more examples of GFP positive roots.

## 2.5. Detection of target mutations

To analyze the mutations or large deletions in the transgenic hairy roots, regions including one or two target sites of gRNAs were PCR amplified using Q5® High-Fidelity DNA polymerase and analyzed by agarose-gel electrophoresis and T7 endonuclease I assay (Fig. 24). If they are wild type and some edited alleles within a root cell population, denaturation and reannealing of amplified target regions will generate amplicons with mismatches (heteroduplexes). These are recognised by T7 endonuclease I and cleaved, creating two smaller fragments that are easily distinguished during agarose-gel electrophoresis.

***AMY2*** was targeted by CRISPR/Cas9 vector **AMY2:gRNA1-gRNA2**. If only AMY2:gRNA2 was functional, it could introduce indels at exon 4 and possibly disrupt severely the open reading frame from that point and onwards. If both AMY2:gRNA1 and AMY2:gRNA2 were functional, the intermediate genomic region could be removed after a double DNA breakage at both target sites. The two genomic regions could still be reconnected by the NHEJ mechanism with a loss of a 2909bp genomic region, including exon 2 (186bp), exon 3 (90bp) and almost half of exon 4 (83bp), if not more. This loss corresponds to 119 amino acid residues. Primers AMY2-intron1_F and AMY2-exon4_R were designed to amplify this target region, but the large amplicon size (3427bp) requires optimization of current cycling conditions. So, we focused on each target site separately and amplified them from AMY2:gRNA1-gRNA2 samples, using primer sets AMY2-intron1_F/R (intron 1) and AMY2-exon4_F/R (exon 4). After purification of PCR products, 250ng of each amplicon were digested with T7 endonuclease I. **Target site 1 (AMY2-gRNA1)** was analyzed from 22

samples. Amplicon size from wild type allele is 566bp and, after digestion with T7 I, two bands at 230bp and 336bp are expected to show. None of them exhibited any sign of T7 I digestion, meaning that these samples were not edited in this site by the CRISPR/Cas9 complex (Fig. 25). **Target site 2 (AMY2-gRNA2)** was analyzed from 23 samples. In this case, amplicon size from wild type allele is 506bp and, after digestion with T7 I, two bands at 218bp and 288bp are expected to show. None of them exhibited any sign of T7 I digestion, meaning that these samples were not edited in this site by the CRISPR/Cas9 complex (Fig. 26). More AMY2:gRNA1-gRNA2 root samples need to be scanned.

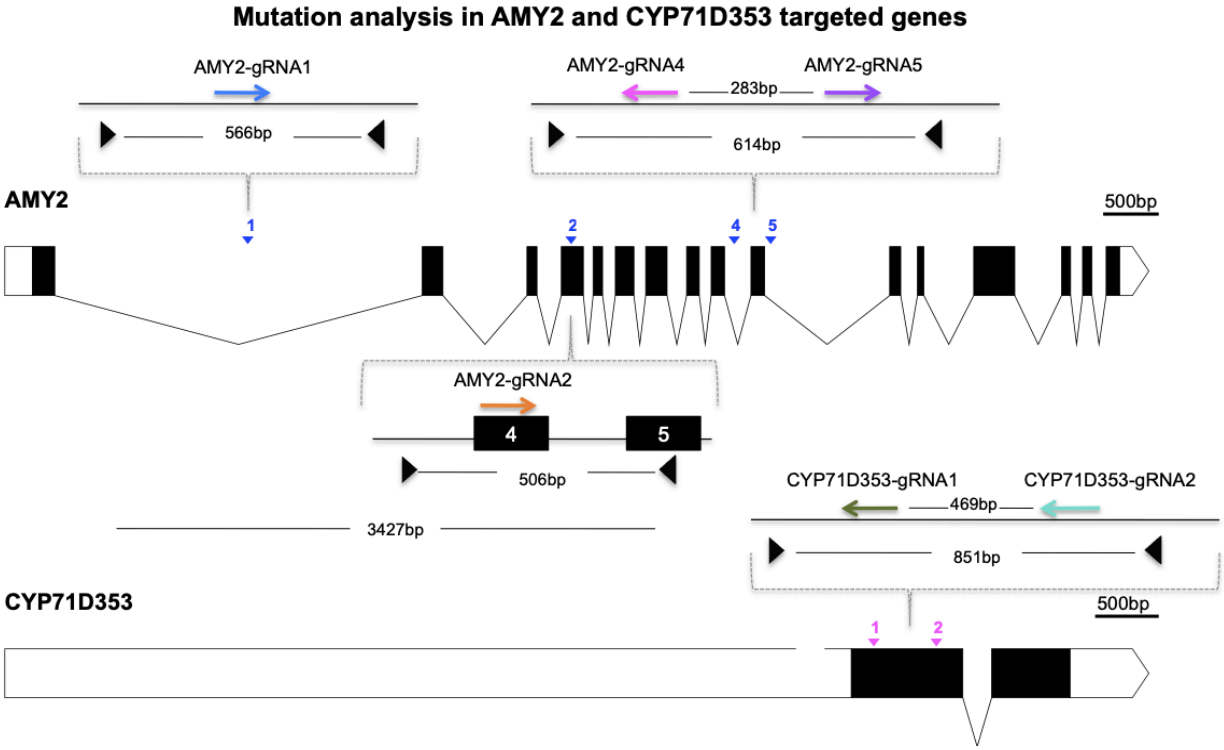



**Figure 24**. **Assessment of mutations at target sites of *AMY2* and *CYP71D353* genes**. Each arrowhead set represents a set of forward and reverse primers used to amplify the targeted region. Amplicon size is also depicted. For amplified regions including two target sites, the distance between target sites is included as well.

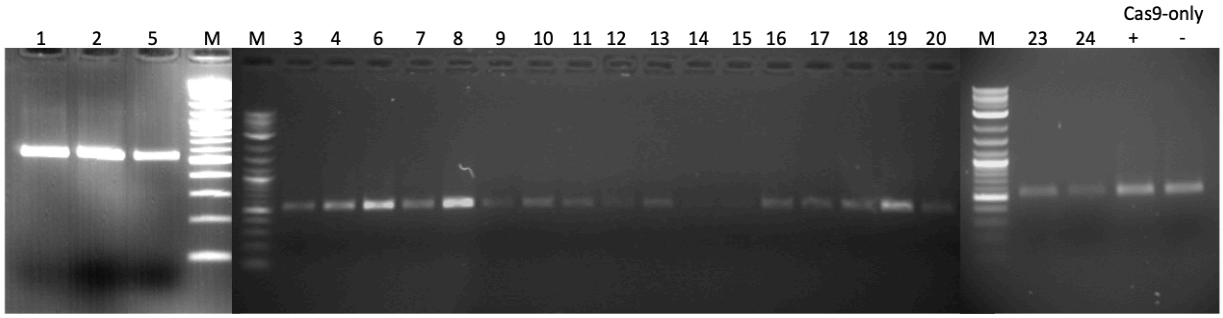



**Figure 25. T7 endonuclease I assay at the amplified *AMY2* intron 1 target region.** Images come from different gel agarose lanes. M: 2-Log DNA Ladder (0.1-10.0 kb), Numbers represent different root samples transformed with the AMY2:gRNA1-gRNA2 T-DNA region. Cas9-only+: MG20 control line digested with T7 I. Cas9-only-: MG20 control line not digested with T7 I.

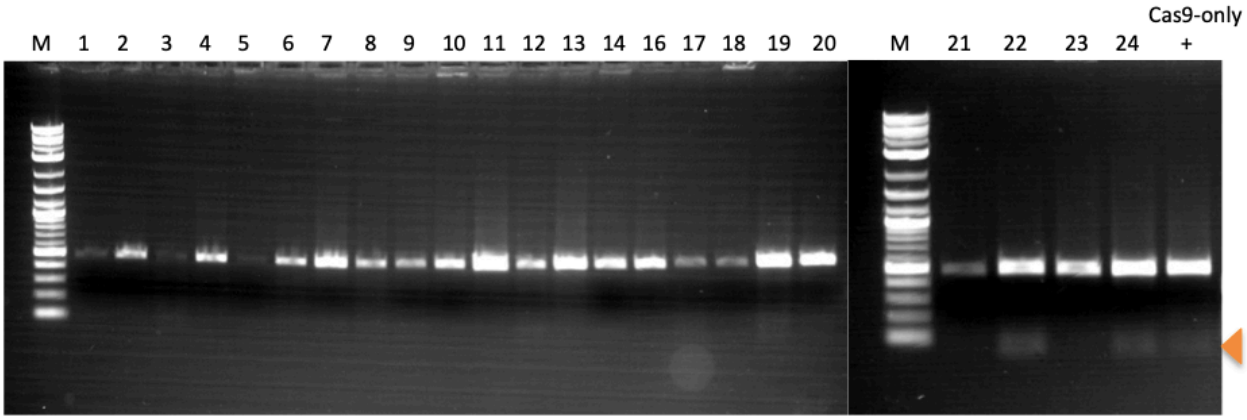



**Figure 26. T7 endonuclease I assay at the amplified *AMY2* exon 4 target region**. Images come from different gel agarose lanes. M: 2-Log DNA Ladder (0.1-10.0 kb), Numbers represent different root samples transformed with the AMY2:gRNA1-gRNA2 T-DNA region. Cas9-only+: MG20 control line digested with T7 I. Orange arrowhead shows primer dimmers.

***AMY2*** was also targeted by CRISPR/Cas9 vector **AMY2:gRNA4-gRNA5**. If both AMY2:gRNA4 and AMY2:gRNA5 were functional, the intermediate genomic region could be removed after a double DNA breakage at both target sites. The two genomic regions could still be reconnected by the NHEJ mechanism with a loss of a 317bp genomic region, including exon 10 (123bp), if not more. This loss corresponds to 49 amino acid residues, including amino acid motif DCTAE that is involved in substrate binding. Primers AMY2-exon10_F and AMY2-exon10_R were designed to amplify this target region that contains two target sites. Gel electrophoresis did not reveal any bands at a lower height (Fig. 27). A faint band over the main product in some samples (1, 2) is a non-specific product. We do not consider it an indication of gene editing. However, if there was any mutation at either or both target sites, it would be detected through T7 I assay. After purification of PCR products, 250ng of each amplicon were digested with T7 endonuclease I. **Target site 4 and 5** were both analyzed simultaneously from 23 samples. Amplicon size from wild type allele is 614bp and cleavage at **target site 4** by T7 I would produce two bands at 107bp and 507bp. Cleavage at **target site 5** would produce two bands at 424bp and 190bp. If **both sites** had mismatches (without an obvious deleted region between them), during T7 endonuclease I assay, two cleavages would produce three bands at 107bp, 317bp and 190bp. None of them exhibited any sign of T7 I digestion, meaning that these samples were not edited in any site by the CRISPR/Cas9 complexes (Fig. 28). More AMY2:gRNA4-gRNA5 root samples need to be scanned.

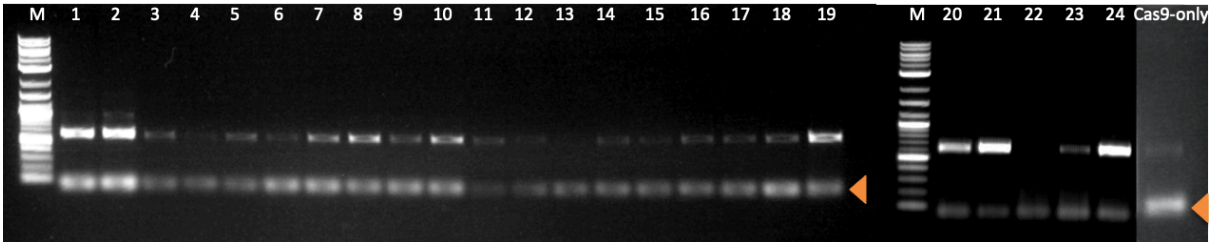



**Figure 27. PCR amplification of *AMY2* target region (exon 10).** Images come from different gel agarose lanes. Amplicon size from wild type allele is 614bp. M: 2-Log DNA Ladder (0.1-10.0 kb), Numbers represent different root samples transformed with the AMY2:gRNA4-gRNA5. Cas9-only: MG20 Cas9-only. Orange arrowheads show primer dimmers.

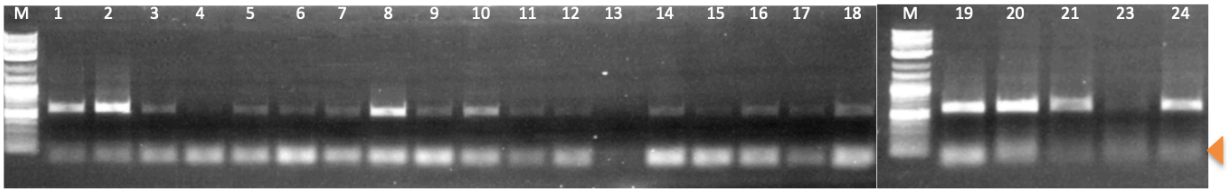



**Figure 28. T7 endonuclease I assay at the amplified *AMY2* exon 10 target region.** Images come from different gel agarose lanes. M: 2-Log DNA Ladder (0.1-10.0 kb), Numbers represent different root samples transformed with the AMY2:gRNA4-gRNA5. Orange arrowhead shows primer dimmers.

Exon 1 of ***CYP71D353*** was targeted by CRISPR/Cas9 vector **CYP71D353:gRNA1-gRNA2.** Either gRNA can introduce indels, or, if both target sites were targeted, the intermediate exonic region of 492bp could be removed. This would translate into a loss of 162 amino acid residues. CYP71D-exon1_F and CYP71D-exon1_R were designed to amplify this target region that contains two target sites. Gel electrophoresis did not reveal any bands at a lower height (Fig. 29). A faint band under the main product is either a reflection or a non-specific product. We do not consider it an indication of gene editing, because it appears in many other samples in the same height. Nevertheless, if there was any mutation at either or both target sites that did not lead to deletion of larger number of nucleotides, it would be detected through T7 I assay. After purification of PCR products, 250ng of each amplicon were digested with T7 endonuclease I. **Target site 1 and 2** were both analyzed simultaneously from 24 samples. Amplicon size from wild type allele is 851bp and cleavage at **target site 1** by T7 I would produce two bands at 180bp and 671bp. Cleavage at **target site 2** would produce two bands at 672bp and 179bp. If **both sites** had mismatches (without an obvious deleted region between them), during T7 endonuclease I assay, two cleavages would produce three bands at 180bp, 492bp and 179bp. None of them exhibited any sign of T7 I digestion, meaning that these samples were not edited in any site by the CRISPR/Cas9 complexes (Fig. 30). More CYP71D353:gRNA1-gRNA2 root samples need to be scanned.

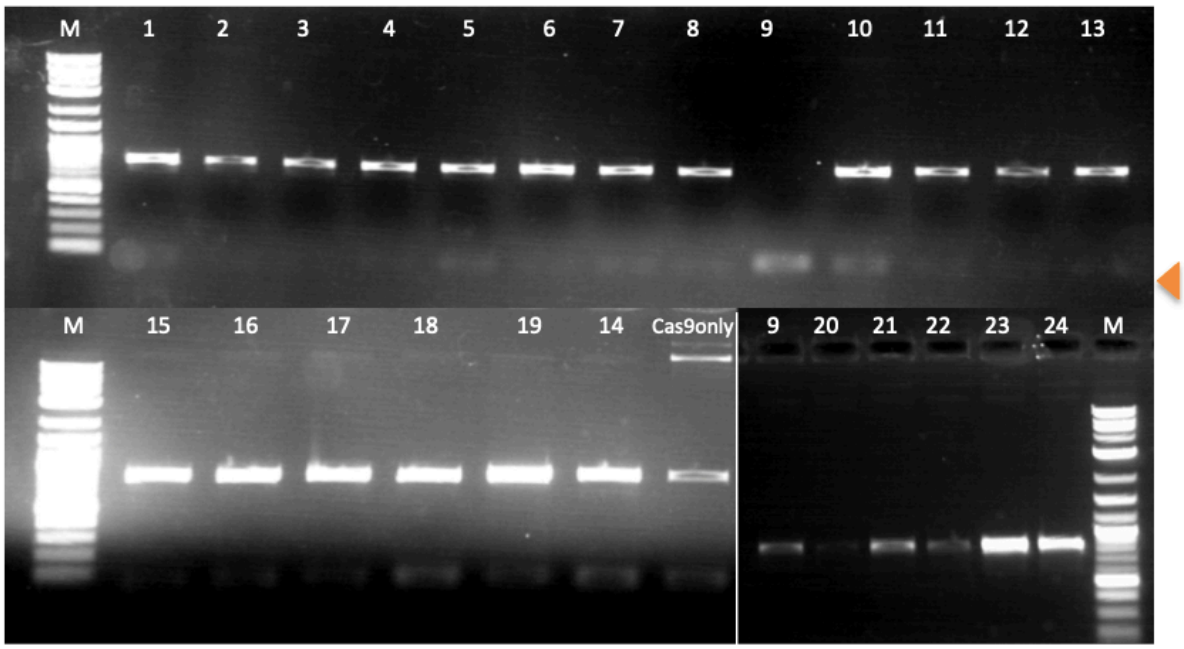



**Figure 29. PCR amplification of *CYP71D353* target region (exon 1).** Images come from different gel agarose lanes. Amplicon size from wild type allele is 851bp. M: 2-Log DNA Ladder (0.1-10.0 kb), Numbers represent different root samples transformed with the CYP71D353:gRNA1-gRNA2. Cas9-only: MG20 Cas9-only. Orange arrowhead shows primer dimmers.

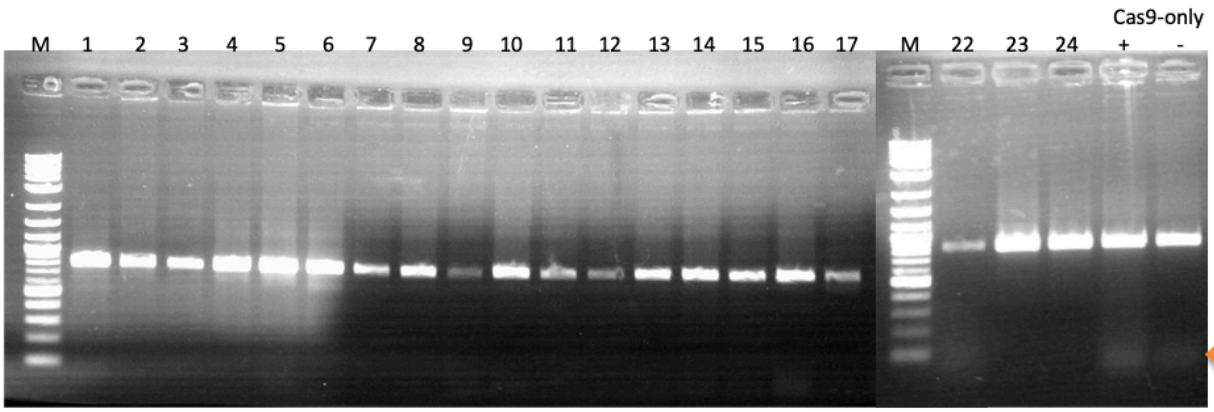



**Figure 30. T7 endonuclease I assay at the amplified *CYP71D353* exon 1 target region.** Images come from different gel agarose lanes. M: 2-Log DNA Ladder (0.1-10.0 kb), Numbers represent different root samples transformed with the CYP71D353:gRNA1-gRNA2 T-DNA region. Cas9-only+: MG20 control line digested with T7 I. Cas9-only-: MG20 control line not digested with T7 I. Orange arrowhead shows primer dimmers.

In the future, we will also analyze the mutation efficacy of the four designed gRNAs against *LSK1*, using the same approach. Amplification of targeted regions was attempted with the appearance of non-specific products. After we have optimised the PCR parameters, we will analyze each target site by T7 I assay.

During this series of hairy root transformations, explants were not subjected to PTT herbicide. Selection of transgenic tissues based on the phenotype after incubation with a herbicide/antibiotic is usually employed during regeneration of calluses, roots and shoots from wild type explants in stable transformation. We decide to also include the selection marker cassette in the CRISPR/Cas9 binary vectors employed in hairy root transformation, in order to verify proper integration of T-DNA region into the plant genome. After confirmation of mutation efficacy during hairy root transformation, the same CRISPR/Cas9 binary vectors will be employed for *A. tumefaciens*-mediated stable transformation.

Regarding off-target effects, we did not analyze any of the predicted off-target sites during the duration of this master thesis. In the future, after confirming on-target mutations, we will also assess off-target mutation frequency. Notably, Modrzejewski et al. (2019) report that mutations are detected in ~3% of analyzed potential off-target sites in plants.

# Discussion-Upcoming experiments

Legumes constitute a highly nutritious source of protein and vegetable oil. Under nitrogen-poor conditions, they are able to participate in symbiotic relationships with bacteria of the genus Rhizobium in exchange for nitrogen fixation. Legumes utilize this nitrogen source to cover their own energy needs, as well as that of the soil environment. Consequently, natural and agricultural ecosystems flourish under their presence (Márquez & J Stougaard, 2005). In addition, legumes, as with all plants, are notorious for their ability to synthesise plethora of specialised metabolites with favourable characteristics towards their adaptation to adverse environmental conditions and their natural growth development. Their secondary metabolism is at the centre of biological research because of its diverse industrial and medicinal applications (Erb & Kliebenstein, 2020). Unfortunately, legume crop yield and quality is affected by numerous factors. Progressive soil chemical and physical degradation, nutrient depletion, overpopulation and prolonged periods of increased temperature and drought are some examples of the adversities that modern society has to deal with. The fields of basic biological and applied research are tasked with finding solutions to improve legume resistance to biotic and abiotic stress factors and increase crop production and quality (Graham & Vance, 2003). To accomplish this, researchers need to understand the plant's physiology and elucidate the molecular pathways of secondary metabolism and symbiotic relationships. *Lotus japonicus* is one of the model legume species employed to uncover the intricate regulatory details behind them (Handberg & Stougaard, 1992).

Investigation of the triterpene biosynthetic metabolism (part of the vast secondary metabolism), nodulation and family of GSK3b/SHAGGY-like kinases revealed an interwoven association between them. The *AMY2* gene cluster and **CYP88D4** seems to regulate different aspects of nodulation, such as infection progression and nodule organogenesis (Krokida et al., 2013; unpublished results). *OSC3* is also involved in nodule organogenesis and *AMY1* expression is deregulated during inoculation with *M. loti* R7A (Delis et al., 2011). *LSK1*, on the other hand, seems to modulate the expression of AON (auto regulation of nodulation) components in response to rhizobium inoculation differently at low and high levels of nitrate independent of its transcription expression levels (Garagounis et al., 2019). Last, in vitro analysis has shown LSK1 binding to lupeol (unpublished results). This triangle of different intertwined biological pathways is worth further probing. Another pair of interlaced pathways is that of triterpene metabolism with *L. japonicus*-FsK interaction. On one hand, *AMY2* expression is induced in FsK-treated plants but on the other hand, FsK seems to restrict the expression of the *AMY2* pathway (unpublished results). It is worthy to explore the putative involvement of other triterpene pathways and *LSK1* in this peculiar symbiosis. Functional analysis of mutants depleted of specific triterpenes or kinase function is the first step towards elucidating these interconnections. We decided to employ a CRISPR/Cas9 system optimised for gene editing in *L. japonicus* and start by generating a series of ***amy2*, *cyp71d353* and** lsk1 **stable silenced plant lines.**

The CRISPR/Cas9 system was assembled as a Golden Gate-based binary vector employed for either *A. rhizogenes*-mediated hairy root transformation or *A. tumefaciens*-mediated stable transformation. First, a CRISPR/Cas9 level 0 collection was created including: promoters LjUbq1 and LjU6 to drive the expression of hSpCas9 and sgRNA respectively; hSpCas9-C'SV40 NLS, turboGFP and Bar coding sequences; CaMV 35S and Actin 2 gene's terminators, and gRNA scaffold. A N' SV40 NLS was utilised as dsoligo with complimentary overhangs to ensure proper translocation of hSpCas9 (Cas9) to the nucleus. At the second phase, a series of CRISPR/Cas9 level 1 transcriptional units was generated based on the order and orientation of the cassettes in the final CRISPR/Cas9 level 2 vector. Apart from the Cas9 expression cassette and the gRNA cassette(s), a reporter gene and selection marker was included in the T-DNA region. A double selection scheme of GFP fluorescence and resistance to herbicide PPT can ensure proper selection of transgenic tissues. Golden Gate-based CRISPR/Cas9 binary vectors have also been generated from other groups for gene editing in different plant species with or without employing additional cloning technologies (e.g. Gateway technology) (Dai et al., 2020; Bernard et al., 2019; Zheng et al., 2020). Each group selected its own cloning strategy to incorporate the Cas9 cassette and gRNA cassette(s) into one vector. The double selection scheme of GFP fluorescence and PPT herbicide has also been utilised before to screen and select for trangenic or transgene-free Arabidopsis plants of the T0, T1 and T2 generations (Zheng et al., 2020). GFP fluorescence or another visible sigh is used to differentiate transformed from non-transformed hairy roots (Kirchner, Niehaus, Debener, Schenk, & Herde, 2017).

Human codon-optimised Cas9 has been used in plants and particularly in *L. japonicus* (Zhang, Yang, Yang, Li, & Guo, 2016; Wang et al., 2016). It was however driven by double CaMV 35S promoter, whose transcriptional activity, and that of single CaMV 35S promoter, is weak in all tissues of *L. japonicus* (Maekawa et al., 2008). This is why we decided to express Cas9 under the LjUbq1 promoter, which has exhibited strong activity in multiple tissues (Maekawa et al., 2008). Regarding the translocation of Cas9 protein to the nucleus, multiple studies employ two NLSs, one at each side (Bernard et al., 2019; Wang et al., 2016). On the other hand, gRNA expression was driven by an endogenous U6 promoter in *L. japonicus* (Wang et al., 2016). We also included the same promoter in our CRISPR/Cas9 system. Many studies encourage the utility of at least two gRNAs to target one gene (Zheng et al., 2020; Wang et al., 2016; Bernard et al., 2019). Last, regarding the architecture of T-DNA region, the orientation of the Cas9 cassette in regard to the one of the gRNA cassette(s) can affect the transcription rate and thus the mutation frequency (Castel, Tomlinson, Locci, Yang, & Jones, 2019). The 'head-to-head' orientation was better than the 'tail-to-head' orientation.

In order to modify and silence *AMY2*, *CYP71D353* and *LSK1*, we designed eleven gRNAs; five targeting *AMY2*, two targeting CYP71D353 and four targeting *LSK1*. We assembled seven CRISPR/Cas9 binary vectors and one Cas9-only, but from these we utilized five, meaning ten out of eleven designed gRNAs were introduced into *L. japonicus* hairy roots. The method of *A. rhizogenes*-mediated hairy root transformation has been used before to test the mutation efficacy of each gRNA in multiple plants, including *L. japonicus* (Zheng et al., 2020; Li et al., 2019; Dai et al., 2020; Wang et al., 2016). After verifying the mutation frequency in hairy roots, the next step is to perform stable transformation, selection of transgenic tissues and regeneration. After some generations, seeds homozygous for a modified allele or alleles can be collected (Wang et al., 2019). During this master thesis we attempted to identify mutations in *AMY2* and *CYP71D353*; biosynthetic genes involved in the triterpene metabolism, and *LSK1*; a signaling kinase involved in nodulation. Until the writing of this manuscript, we analysed ~23 samples from AMY2:gRNA1-gRNA2, AMY2:gRNA4-gRNA5 and CYP71D353:gRNA1-gRNA2 hairy roots and were unable to detect any gene editing effect. None of the LSK1:gRNA1-gRNA3 and LSK1:gRNA2-gRNA4 hairy root samples were analysed due to technical problems. In the upcoming weeks, we plant to analyse them, as well as the rest of the other hairy root samples. The fact that there are four published works recording the use of CRISPR/Cas9 in *Lotus japonicus* with many CRISPR/Cas9 regulatory parts and methodologies common between our CRISPR/Cas9 systems gives us confidence that we will identify mutations from at least one gRNA targeting each gene.

The first gene editing event to be recorded was in *LjSYMR*K (symbiosis receptor-like kinase) through *A. tumefaciens*-mediated stable transformation. Wang and his group achieved ~35% mutagenic efficiency in 20 T0 transgenic plants, with two of them containing biallelic homozygous mutations with a 2-bp deletion near the PAM region. Three more genes were modified by two gRNAs, *LjLb1*, *LjLb2* and *LjLb3*. 20 out of 70 hairy root transgenic plants exhibited white nodules, with at least two *LjLbs* disrupted in each plant. A triple mutant knockout of LjLbs was also obtained by stable transformation (Wang et al., 2016). This mutant was further used in functional analysis (Wang et al., 2019). Another gene to be modified was *LjCYP716A51*, a triterpene biosynthetic gene that metabolizes lupeol to betulinic acid. It was targeted by two gRNAs and modified in transgenic hairy roots (Suzuki et al., 2019). Finally, *LjCZF1* and *LjCZF2*, two homologous C3HC4-type RING finger proteins, were targeted by two gRNAs each. Both were disrupted in trangenic hairy roots. Silencing of *LjCZF1* greatly affected nodulation (Cai et al., 2018). Analysis of off-target effects is not mentioned in any of the published works regarding *L. japonicus.* We plant to assess off-target mutation frequency after confirming the on-target one.

In conclusion, during this master thesis we created a Golden Gate-based cloning strategy for CRISPR/Cas9-mediated genome editing in *L. japonicus*. We assembled several CRISPR/Cas9 binary vector targeting three genes and introduced them into *A. rhizogenes*. We next proceeded with assessing the mutation efficacy of each gRNA in transgenic hairy roots. Due to the low number of analyzed samples, we were unable to detect any gene editing effect until the writing of this manuscript. In the upcoming weeks, we will analyse the rest of the hairy root samples and employ the other two CRISPR/Cas9 binary vectors that were not included in this batch of hairy root transformations. After confirming which gRNAs are functional, we will proceed with the *A. tumefaciens*-mediated stable transformation to acquire *amy2*, *cyp71d353* and *lsk1* mutant lines. In the future, we will also target *OSC3* and *AMY1*, as well both *AMY1* and *AMY2*. Our aim is to utilize the CRISPR/Cas9-mediated mutant seeds for functional analysis and understand better the interconnection between triterpene metabolism, symbioses and *LSK1*.

# Bibliography


1. Abe, I. (2007). Enzymatic synthesis of cyclic triterpenes. *Natural Product Reports*, *24*(6), 1311. https://doi.org/10.1039/b616857b
2. Abe, Ikuro., Rohmer, Michel., & Prestwich, G. D. (1993). Enzymatic cyclization of squalene and oxidosqualene to sterols and triterpenes. *Chemical Reviews*, *93*(6), 2189–2206. https://doi.org/10.1021/cr00022a009
3. Antolín-Llovera, M., Ried, Martina K., & Parniske, M. (2014). Cleavage of the SYMBIOSIS RECEPTOR-LIKE KINASE Ectodomain Promotes Complex Formation with Nod Factor Receptor 5. *Current Biology*, *24*(4), 422–427. https://doi.org/10.1016/j.cub.2013.12.053
4. Badhan, S., Ball, A. S., & Mantri, N. (2021). First Report of CRISPR/Cas9 Mediated DNA-Free Editing of 4CL and RVE7 Genes in Chickpea Protoplasts. *International Journal of Molecular Sciences*, *22*(1), 396. https://doi.org/10.3390/ijms22010396
5. Bahramnejad, B., Naji, M., Bose, R., & Jha, S. (2019). A critical review on use of Agrobacterium rhizogenes and their associated binary vectors for plant transformation. *Biotechnology Advances*, *37*(7), 107405. https://doi.org/10.1016/j.biotechadv.2019.06.004
6. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., … Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science (New York, N.Y.)*, *315*(5819), 1709–1712. https://doi.org/10.1126/science.1138140
7. Barrangou, R., & Marraffini, Luciano A. (2014). CRISPR-Cas Systems: Prokaryotes Upgrade to Adaptive Immunity. *Molecular Cell*, *54*(2), 234–244. https://doi.org/10.1016/j.molcel.2014.03.011
8. Belhaj, K., Chaparro-Garcia, A., Kamoun, S., & Nekrasov, V. (2013). Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. *Plant Methods*, *9*(1), 39. https://doi.org/10.1186/1746-4811-9-39
9. Belhaj, K., Chaparro-Garcia, A., Kamoun, S., Patron, N. J., & Nekrasov, V. (2015). Editing plant genomes with CRISPR/Cas9. *Current Opinion in Biotechnology*, *32*, 76–84. https://doi.org/10.1016/j.copbio.2014.11.007
10. Bernard, G., Gagneul, D., Alves Dos Santos, H., Etienne, A., Hilbert, J.-L., & Rambaud, C. (2019). Efficient Genome Editing Using CRISPR/Cas9 Technology in Chicory. *International Journal of Molecular Sciences*, *20*(5), 1155. https://doi.org/10.3390/ijms20051155
11. Bibikova, M., Golic, M., Golic, K. G., & Carroll, D. (2002). Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. *Genetics*, *161*(3), 1169–1175. Retrieved from https://pubmed.ncbi.nlm.nih.gov/12136019/
12. Bolotin, A., Quinquis, B., Sorokin, A., & Ehrlich, S. D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, *151*(8), 2551–2561. https://doi.org/10.1099/mic.0.28048-0
13. Boogerd, F. C., & van Rossum, D. (1997). Nodulation of groundnut by Bradyrhizobium: a simple infection process by crack entry. *FEMS Microbiology Reviews*, *21*(1), 5–27. https://doi.org/10.1111/j.1574-6976.1997.tb00342.x
14. Brouns, S. J. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J. H., Snijders, A. P. L., … van der Oost, J. (2008). Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science*, *321*(5891), 960–964. https://doi.org/10.1126/science.1159689
15. Cai, K., Yin, J., Chao, H., Ren, Y., Jin, L., Cao, Y., … Zhang, Z. (2018). A C3HC4-type RING finger protein regulates rhizobial infection and nodule organogenesis in Lotus japonicus. *Journal of Integrative Plant Biology*, *60*(9), 878–896. https://doi.org/10.1111/jipb.12703
16. Cai, Y., Chen, L., Liu, X., Sun, S., Wu, C., Jiang, B., … Hou, W. (2015). CRISPR/Cas9-Mediated Genome Editing in Soybean Hairy Roots. *PLOS ONE*, *10*(8), e0136064. https://doi.org/10.1371/journal.pone.0136064
17. Castel, B., Tomlinson, L., Locci, F., Yang, Y., & Jones, J. D. G. (2019). Optimization of T-DNA architecture for Cas9-mediated mutagenesis in Arabidopsis. *PLOS ONE*, *14*(1), e0204778. https://doi.org/10.1371/journal.pone.0204778
18. Charpentier, M., Bredemeier, R., Wanner, G., Takeda, N., Schleiff, E., & Parniske, M. (2008). Lotus japonicus CASTOR and POLLUX are ion channels essential for perinuclear calcium spiking in legume root endosymbiosis. *The Plant Cell*, *20*(12), 3467–3479. https://doi.org/10.1105/tpc.108.063255
19. Che, P., Chang, S., Simon, M. K., Zhang, Z., Shaharyar, A., Ourada, J., … Jones, T. J. (2021). Developing a rapid and highly efficient cowpea regeneration, transformation and genome editing system using embryonic axis explants. *The Plant Journal: For Cell and Molecular Biology*, *106*(3), 817–830. https://doi.org/10.1111/tpj.15202
20. Chung, S. Y., Seki, H., Fujisawa, Y., Shimoda, Y., Hiraga, S., Nomura, Y., … Muranaka, T. (2020). A cellulose synthase-derived enzyme catalyses 3- O -glucuronosylation in saponin biosynthesis. *Nature Communications*, *11*(1), 5664. https://doi.org/10.1038/s41467-020-19399-0
21. Claisse, G., Charrier, B., & Kreis, M. (2007). The Arabidopsis thaliana GSK3/Shaggy like kinase AtSK3-2 modulates

floral cell expansion. *Plant Molecular Biology*, *64*(1-2), 113–124. https://doi.org/10.1007/s11103-007-9138-y
22. Coleman, J. J. (2015). The Fusarium solani species complex: ubiquitous pathogens of agricultural importance. *Molecular Plant Pathology*, *17*(2), 146–158. https://doi.org/https://doi.org/10.1111/mpp.12289
23. Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., … Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*, *339*(6121), 819–823. https://doi.org/10.1126/science.1231143
24. Curtin, S. J., Xiong, Y., Michno, J.-M., Campbell, B. W., Stec, A. O., Čermák, T., … Stupar, R. M. (2017). CRISPR/Cas9 and TALENs generate heritable mutations for genes involved in small RNA processing ofGlycine maxandMedicago truncatula. *Plant Biotechnology Journal*, *16*(6), 1125–1137. https://doi.org/10.1111/pbi.12857
25. Dai, Y., Hu, G., Dupas, A., Medina, L., Blandels, N., San Clemente, H., … Cassan-Wang, H. (2020). Implementing the CRISPR/Cas9 Technology in Eucalyptus Hairy Roots Using Wood-Related Genes. *International Journal of Molecular Sciences*, *21*(10). https://doi.org/10.3390/ijms21103408
26. De Rybel, B., Audenaert, D., Vert, G., Rozhon, W., Mayerhofer, J., Peelman, F., … Beeckman, T. (2009). Chemical inhibition of a subset of Arabidopsis thaliana GSK3-like kinases activates brassinosteroid signaling. *Chemistry & Biology*, *16*(6), 594–604. https://doi.org/10.1016/j.chembiol.2009.04.008
27. Delis, C., Krokida, A., Georgiou, S., Peña-Rodríguez, L. M., Kavroulakis, N., Ioannou, E., … Papadopoulou, K. K. (2011). Role of lupeol synthase in Lotus japonicus nodule formation. *The New Phytologist*, *189*(1), 335–346. https://doi.org/10.1111/j.1469-8137.2010.03463.x
28. Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., … Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, *471*(7340), 602–607. https://doi.org/10.1038/nature09886
29. Denarie, J., & Cullimore, J. (1993). Lipo-oligosaccharide nodulation factors: A new class of signaling molecules mediating recognition and morphogenesis. *Cell*, *74*(6), 951–954. https://doi.org/10.1016/0092-8674(93)90717-5
30. Engler, C., Youles, M., Gruetzner, R., Ehnert, T.-M., Werner, S., Jones, J. D. G., … Marillonnet, S. (2014). A Golden Gate Modular Cloning Toolbox for Plants. *ACS Synthetic Biology*, *3*(11), 839–843. https://doi.org/10.1021/sb4001504
31. Erb, M., & Kliebenstein, D. J. (2020). Plant Secondary Metabolites as Defenses, Regulators, and Primary Metabolites: The Blurred Functional Trichotomy. *Plant Physiology*, *184*(1), 39–52. https://doi.org/10.1104/pp.20.00433
32. Feng, Z., Zhang, B., Ding, W., Liu, X., Yang, D.-L., Wei, P., … Zhu, J.-K. (2013). Efficient genome editing in plants using a CRISPR/Cas system. *Cell Research*, *23*(10), 1229–1232. https://doi.org/10.1038/cr.2013.114
33. Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., & Sander, J. D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology*, *31*(9), 822–826. https://doi.org/10.1038/nbt.2623
34. Gage, D. J., & Margolin, W. (2000). Hanging by a thread: invasion of legume plants by rhizobia. *Current Opinion in Microbiology*, *3*(6), 613–617. https://doi.org/10.1016/s1369-5274(00)00149-1
35. Gao, Z., Harwig, A., Berkhout, B., & Herrera-Carrillo, E. (2017). Mutation of nucleotides around the +1 position of type 3 polymerase III promoters: The effect on transcriptional activity and start site usage. *Transcription*, *8*(5), 275–287. https://doi.org/10.1080/21541264.2017.1322170
36. Gao, Z., Herrera-Carrillo, E., & Berkhout, B. (2018). Delineation of the Exact Transcription Termination Signal for Type 3 Polymerase III. *Molecular Therapy. Nucleic Acids*, *10*(2), 36–44. https://doi.org/10.1016/j.omtn.2017.11.006
37. Garagounis, C., Tsikou, D., Plitsi, P. K., Psarrakou, I. S., Avramidou, M., Stedel, C., … Papadopoulou, K. K. (2019). Lotus SHAGGY-like kinase 1 is required to suppress nodulation in Lotus japonicus. *The Plant Journal*, *98*(2), 228–242. https://doi.org/10.1111/tpj.14207
38. Garantonakis, N., Pappas, M. L., Varikou, K., Skiada, V., Broufas, G. D., Kavroulakis, N., & Papadopoulou, K. K. (2018). Tomato Inoculation With the Endophytic Strain Fusarium solani K Results in Reduced Feeding Damage by the Zoophytophagous Predator Nesidiocoris tenuis. *Frontiers in Ecology and Evolution*, *6*. https://doi.org/10.3389/fevo.2018.00126
39. Garneau, J. E., Dupuis, M.-È., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., … Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, *468*(7320), 67–71. https://doi.org/10.1038/nature09523
40. Gasparis, S., Kała, M., Przyborowski, M., Łyżnik, L. A., Orczyk, W., & Nadolska-Orczyk, A. (2018). A simple and efficient CRISPR/Cas9 platform for induction of single and multiple, heritable mutations in barley (Hordeum vulgare L.). *Plant Methods*, *14*(1). https://doi.org/10.1186/s13007-018-0382-8
41. Graham, P. H., & Vance, C. P. (2003). Legumes: Importance and Constraints to Greater Use. *Plant Physiology*, *131*(3), 872–877. https://doi.org/10.1104/pp.017004
42. Groth, M., Takeda, N., Perry, J., Uchida, H., Dräxl, S., Brachmann, A., … Parniske, M. (2010). NENA, a Lotus japonicus homolog of Sec13, is required for rhizodermal infection by arbuscular mycorrhiza fungi and rhizobia but dispensable for cortical endosymbiotic development. *The Plant Cell*, *22*(7), 2509–2526. https://doi.org/10.1105/tpc.109.069807
43. Guinel, F. C., & Geil, R. D. (2002). A model for the development of the rhizobial and arbuscular mycorrhizal symbioses in legumes and its use to understand the roles of ethylene in the establishment of these two symbioses.

*Canadian Journal of Botany*, *80*(7), 695–720. https://doi.org/10.1139/b02-066
44. Handberg, K., & Stougaard, J. (1992). Lotus japonicus, an autogamous, diploid legume species for classical and molecular genetics. *The Plant Journal*, *2*(4), 487–496. https://doi.org/10.1111/j.1365-313x.1992.00487.x
45. Hashimoto, R., Ueta, R., Abe, C., Osakabe, Y., & Osakabe, K. (2018). Efficient Multiplex Genome Editing Induces Precise, and Self-Ligated Type Mutations in Tomato Plants. *Frontiers in Plant Science*, *9*(), 916. https://doi.org/10.3389/fpls.2018.00916
46. Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., … Zhang, F. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology*, *31*(9), 827–832. https://doi.org/10.1038/nbt.2647
47. Imaizumi-Anraku, H., Takeda, N., Charpentier, M., Perry, J., Miwa, H., Umehara, Y., … Kawaguchi, M. (2005). Plastid proteins crucial for symbiotic fungal and bacterial entry into plant roots. *Nature*, *433*(7025), 527–531. https://doi.org/10.1038/nature03237
48. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., & Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *Journal of Bacteriology*, *169*(12), 5429–5433. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC213968/
49. Iturbe-Ormaetxe, I., Haralampidis, K., Papadopoulou, K., & Osbourn, A. E. (2003). Molecular cloning and characterization of triterpene synthases from Medicago truncatula and Lotus japonicus. *Plant Molecular Biology*, *51*(5), 731–743. https://doi.org/10.1023/a:1022519709298
50. Jacobs, T. B., LaFayette, P. R., Schmitz, R. J., & Parrott, W. A. (2015). Targeted genome modifications in soybean with CRISPR/Cas9. *BMC Biotechnology*, *15*(1). https://doi.org/10.1186/s12896-015-0131-2
51. Jansen, Ruud., Embden, Jan. D. A. van, Gaastra, Wim., & Schouls, Leo. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology*, *43*(6), 1565–1575. https://doi.org/10.1046/j.1365-2958.2002.02839.x
52. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, *337*(6096), 816–821. https://doi.org/10.1126/science.1225829
53. Jonak, C., & Hirt, H. (2002). Glycogen synthase kinase 3/SHAGGY-like kinases in plants: an emerging family with novel functions. *Trends in Plant Science*, *7*(10), 457–461. https://doi.org/10.1016/s1360-1385(02)02331-2
54. Kanamori, N., Madsen, L. H., Radutoiu, S., Frantescu, M., Quistgaard, E. M. H., Miwa, H., … Stougaard, J. (2006). A nucleoporin is required for induction of Ca2+ spiking in legume nodule development and essential for rhizobial and fungal symbiosis. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(2), 359–364. https://doi.org/10.1073/pnas.0508883103
55. Kavroulakis, N., Doupis, G., Papadakis, I. E., Ehaliotis, C., & Papadopoulou, K. K. (2018). Tolerance of tomato plants to water stress is improved by the root endophyte Fusarium solani FsK. *Rhizosphere*, *6*, 77–85. https://doi.org/10.1016/j.rhisph.2018.04.003
56. Kavroulakis, N., Ntougias, S., Zervakis, G. I., Ehaliotis, C., Haralampidis, K., & Papadopoulou, K. K. (2007). Role of ethylene in the protection of tomato plants against soil-borne fungal pathogens conferred by an endophytic Fusarium solani strain. *Journal of Experimental Botany*, *58*(14), 3853–3864. https://doi.org/10.1093/jxb/erm230
57. Kemen, A. C., Honkanen, S., Melton, R. E., Findlay, K. C., Mugford, S. T., Hayashi, K., … Osbourn, A. (2014). Investigation of triterpene synthesis and regulation in oats reveals a role for β-amyrin in determining root epidermal cell patterning. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(23), 8679–8684. https://doi.org/10.1073/pnas.1401553111
58. Kirchner, T. W., Niehaus, M., Debener, T., Schenk, M. K., & Herde, M. (2017). Efficient generation of mutations mediated by CRISPR/Cas9 in the hairy root transformation system of Brassica carinata. *PLOS ONE*, *12*(9), e0185429. https://doi.org/10.1371/journal.pone.0185429
59. Krokida, A., Delis, C., Geisler, K., Garagounis, C., Tsikou, D., Peña-Rodríguez, L. M., … Papadopoulou, K. K. (2013). A metabolic gene cluster in Lotus japonicus discloses novel enzyme functions and products in triterpene biosynthesis. *The New Phytologist*, *200*(3), 675–690. https://doi.org/10.1111/nph.12414
60. Kumar, V., & Jain, M. (2014). The CRISPR–Cas system for plant genome editing: advances and opportunities. *Journal of Experimental Botany*, *66*(1), 47–57. https://doi.org/10.1093/jxb/eru429
61. Laszczyk, M. N. (2009). Pentacyclic triterpenes of the lupane, oleanane and ursane group as tools in cancer therapy. *Planta Medica*, *75*(15), 1549–1560. https://doi.org/10.1055/s-0029-1186102
62. Li, C., Nguyen, V., Liu, J., Fu, W., Chen, C., Yu, K., & Cui, Y. (2019). Mutagenesis of seed storage protein genes in Soybean using CRISPR/Cas9. *BMC Research Notes*, *12*(1). https://doi.org/10.1186/s13104-019-4207-2
63. Li, J.-F., Norville, J. E., Aach, J., McCormack, M., Zhang, D., Bush, J., … Sheen, J. (2013). Multiplex and homologous recombination–mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. *Nature Biotechnology*, *31*(8), 688–691. https://doi.org/10.1038/nbt.2654
64. Li, T., Huang, S., Zhao, X., Wright, D. A., Carpenter, S., Spalding, M. H., … Yang, B. (2011). Modularly assembled

designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Research*, *39*(14), 6315–6325. https://doi.org/10.1093/nar/gkr188
65. Liang, Z., Zhang, K., Chen, K., & Gao, C. (2014). Targeted Mutagenesis in Zea mays Using TALENs and the CRISPR/Cas System. *Journal of Genetics and Genomics*, *41*(2), 63–68. https://doi.org/10.1016/j.jgg.2013.12.001
66. Liang, Z., Zong, Y., & Gao, C. (2016). An Efficient Targeted Mutagenesis System Using CRISPR/Cas in Monocotyledons. *Current Protocols in Plant Biology*, *1*(2), 329–344. https://doi.org/10.1002/cppb.20021
67. Limpens, E., Franken, C., Smit, P., Willemse, J., Bisseling, T., & Geurts, R. (2003). LysM Domain Receptor Kinases Regulating Rhizobial Nod Factor-Induced Infection. *Science*, *302*(5645), 630–633. https://doi.org/10.1126/science.1090074
68. Liu, H., Ding, Y., Zhou, Y., Jin, W., Xie, K., & Chen, L.-L. (2017). CRISPR-P 2.0: An Improved CRISPR-Cas9 Tool for Genome Editing in Plants. *Molecular Plant*, *10*(3), 530–532. https://doi.org/10.1016/j.molp.2017.01.003
69. Ma, X., Chen, L., Zhu, Q., Chen, Y., & Liu, Y.-G. (2015). Rapid Decoding of Sequence-Specific Nuclease-Induced Heterozygous and Biallelic Mutations by Direct Sequencing of PCR Products. *Molecular Plant*, *8*(8), 1285–1287. https://doi.org/10.1016/j.molp.2015.02.012
70. Madsen, E. B., Madsen, L. H., Radutoiu, S., Olbryt, M., Rakwalska, M., Szczyglowski, K., … Stougaard, J. (2003). A receptor kinase gene of the LysM type is involved in legumeperception of rhizobial signals. *Nature*, *425*(6958), 637–640. https://doi.org/10.1038/nature02045
71. Maekawa, T., Kusakabe, M., Shimoda, Y., Sato, S., Tabata, S., Murooka, Y., & Hayashi, M. (2008). Polyubiquitin promoter-based binary vectors for overexpression and gene silencing in Lotus japonicus. *Molecular Plant-Microbe Interactions: MPMI*, *21*(4), 375–382. https://doi.org/10.1094/MPMI-21-4-0375
72. Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J. J., Charpentier, E., Horvath, P., … Koonin, E. V. (2011). Evolution and classification of the CRISPR–Cas systems. *Nature Reviews Microbiology*, *9*(6), 467–477. https://doi.org/10.1038/nrmicro2577
73. Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., … Church, G. M. (2013). RNA-Guided Human Genome Engineering via Cas9. *Science*, *339*(6121), 823–826. https://doi.org/10.1126/science.1232033
74. Márquez, A. J., & J Stougaard. (2005). *Lotus japonicus handbook*. Dordrecht: Springer.
75. Marraffini, L. A., & Sontheimer, E. J. (2008). CRISPR Interference Limits Horizontal Gene Transfer in Staphylococci by Targeting DNA. *Science*, *322*(5909), 1843–1845. https://doi.org/10.1126/science.1165771
76. Marraffini, L. A., & Sontheimer, E. J. (2010). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nature Reviews Genetics*, *11*(3), 181–190. https://doi.org/10.1038/nrg2749
77. Martín-Pizarro, C., Triviño, J. C., & Posé, D. (2018). Functional analysis of the TM6 MADS-box gene in the octoploid strawberry by CRISPR/Cas9-directed mutagenesis. *Journal of Experimental Botany*, *70*(3), 885–895. https://doi.org/10.1093/jxb/ery400
78. Modrzejewski, D., Hartung, F., Sprink, T., Krause, D., Kohl, C., & Wilhelm, R. (2019). What is the available evidence for the range of applications of genome-editing as a new tool for plant trait modification and the potential occurrence of associated off-target effects: a systematic map. *Environmental Evidence*, *8*(1). https://doi.org/10.1186/s13750-019-0171-5
79. Mojica, F. J. M., Diez-Villasenor, C., Garcia-Martinez, J., & Soria, E. (2005). Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *Journal of Molecular Evolution*, *60*(2), 174–182. https://doi.org/10.1007/s00239-004-0046-3
80. Moses, T., Pollier, J., Thevelein, J. M., & Goossens, A. (2013). Bioengineering of plant (tri)terpenoids: from metabolic engineering of plants to synthetic biologyin vivoandin vitro. *New Phytologist*, *200*(1), 27–43. https://doi.org/10.1111/nph.12325
81. Mun, T., Bachmann, A., Gupta, V., Stougaard, J., & Andersen, S. U. (2016). Lotus Base: An integrated information portal for the model legume Lotus japonicus. *Scientific Reports*, *6*(1). https://doi.org/10.1038/srep39447
82. Naito, Y., Hino, K., Bono, H., & Ui-Tei, K. (2014). CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. *Bioinformatics*, *31*(7), 1120–1123. https://doi.org/10.1093/bioinformatics/btu743
83. Nakayasu, M., Akiyama, R., Lee, H. J., Osakabe, K., Osakabe, Y., Watanabe, B., … Mizutani, M. (2018). Generation of α-solanine-free hairy roots of potato by CRISPR/Cas9 mediated genome editing of the St16DOX gene. *Plant Physiology and Biochemistry: PPB*, *131*, 70–77. https://doi.org/10.1016/j.plaphy.2018.04.026
84. Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J. D. G., & Kamoun, S. (2013). Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. *Nature Biotechnology*, *31*(8), 691–693. https://doi.org/10.1038/nbt.2655
85. Oldroyd, G. E. D. (2013). Speak, friend, and enter: signaling systems that promote beneficial symbiotic associations in plants. *Nature Reviews. Microbiology*, *11*(4), 252–263. https://doi.org/10.1038/nrmicro2990
86. Oldroyd, G. E. D., & Downie, J. A. (2008). Coordinating Nodule Morphogenesis with Rhizobial Infection in Legumes. *Annual Review of Plant Biology*, *59*(1), 519–546. https://doi.org/10.1146/annurev.arplant.59.032607.092839
87. Oldroyd, G. E. D., Murray, J. D., Poole, P. S., & Downie, J. A. (2011). The rules of engagement in the legume-

rhizobial symbiosis. *Annual Review of Genetics*, *45*(45), 119–144. https://doi.org/10.1146/annurev-genet-110410-132549
88. Papadopoulou, K., Melton, R. E., Leggett, M., Daniels, M. J., & Osbourn, A. E. (1999). Compromised disease resistance in saponin-deficient plants. *Proceedings of the National Academy of Sciences*, *96*(22), 12923–12928. https://doi.org/10.1073/pnas.96.22.12923
89. Pappas, M. L., Liapoura, M., Papantoniou, D., Avramidou, M., Kavroulakis, N., Weinhold, A., … Papadopoulou, K. K. (2018). The Beneficial Endophytic Fungus Fusariumsolani Strain K Alters Tomato Responses Against Spider Mites to the Benefit of the Plant. *Frontiers in Plant Science*, *9*(1603). https://doi.org/10.3389/fpls.2018.01603
90. Pattanayak, V., Lin, S., Guilinger, J. P., Ma, E., Doudna, J. A., & Liu, D. R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature Biotechnology*, *31*(9), 839–843. https://doi.org/10.1038/nbt.2673
91. Peters, N. K., Frost, J. W., & Long, S. R. (1986). A plant flavone, luteolin, induces expression of Rhizobium meliloti nodulation genes. *Science (New York, N.Y.)*, *233*(4767), 977–980. https://doi.org/10.1126/science.3738520
92. Pourcel, C., Salvigno, G., & Vergnaud, G. (2005). CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, *151*(3), 653–663. https://doi.org/10.1099/mic.0.27437-0
93. Puchta, H. (2004). The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. *Journal of Experimental Botany*, *56*(409), 1–14. https://doi.org/10.1093/jxb/eri025
94. Puchta, H., Dujon, B., & Hohn, B. (1996). Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(10), 5055–5060. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC39405/
95. Radutoiu, S., Madsen, L. H., Madsen, E. B., Felle, H. H., Umehara, Y., Grønlund, M., … Stougaard, J. (2003). Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature*, *425*(6958), 585–592. https://doi.org/10.1038/nature02039
96. Redmond, J. W., Batley, M., Djordjevic, M. A., Innes, R. W., Kuempel, P. L., & Rolfe, B. G. (1986). Flavones induce expression of nodulation genes in Rhizobium. *Nature*, *323*(6089), 632–635. https://doi.org/10.1038/323632a0
97. Ron, M., Kajala, K., Pauluzzi, G., Wang, D., Reynoso, M. A., Zumstein, K., … Brady, S. M. (2014). Hairy Root Transformation Using Agrobacterium rhizogenes as a Tool for Exploring Cell Type-Specific Gene Expression and Function Using Tomato as a Model. *PLANT PHYSIOLOGY*, *166*(2), 455–469. https://doi.org/10.1104/pp.114.239392
98. Roychowdhury, D., Halder, M., & Jha, S. (2017). Agrobacterium rhizogenes-Mediated Transformation in Medicinal Plants: Genetic Stability in Long-Term Culture (S. Jha, Ed.). Retrieved June 22, 2021, from Springer Link website: https://link.springer.com/content/pdf/10.1007%2F978-3-319-28669-3_8.pdf
99. Saito, K., Yoshikawa, M., Yano, K., Miwa, H., Uchida, H., Asamizu, E., … Kawaguchi, M. (2007). NUCLEOPORIN85 Is Required for Calcium Spiking, Fungal and Bacterial Symbioses, and Seed Production in Lotus japonicus. *The Plant Cell*, *19*(2), 610–624. https://doi.org/10.1105/tpc.106.046938
100. Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., & Siksnys, V. (2011). The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic Acids Research*, *39*(21), 9275–9282. https://doi.org/10.1093/nar/gkr606
101. Sawai, S. (2006). Functional and structural analysis of genes encoding oxidosqualene cyclases of Lotus japonicus. *Plant Science*, *170*(2), 247–257. https://doi.org/10.1016/j.plantsci.2005.08.027
102. Shan, S., Soltis, P. S., Soltis, D. E., & Yang, B. (2020). Considerations in adapting CRISPR/Cas9 in nongenetic model plant systems. *Applications in Plant Sciences*, *8*(1). https://doi.org/10.1002/aps3.11314
103. Singh, S., Katzer, K., Lambert, J., Cerri, M., & Parniske, M. (2014). CYCLOPS, A DNA-Binding Transcriptional Activator, Orchestrates Symbiotic Root Nodule Development. *Cell Host & Microbe*, *15*(2), 139–152. https://doi.org/10.1016/j.chom.2014.01.011
104. Skiada, V., Avramidou, M., Bonfante, P., Genre, A., & Papadopoulou, K. K. (2020). An endophytic Fusarium –legume association is partially dependent on the common symbiotic signaling pathway. *New Phytologist*, *226*(5), 1429–1444. https://doi.org/10.1111/nph.16457
105. Skiada, V., Faccio, A., Kavroulakis, N., Genre, A., Bonfante, P., & Papadopoulou, K. K. (2019). Colonization of legumes by an endophytic Fusarium solani strain FsK reveals common features to symbionts or pathogens. *Fungal Genetics and Biology*, *127*, 60–74. https://doi.org/10.1016/j.fgb.2019.03.003
106. Smil, V. (1999). Nitrogen in crop production: An account of global flows. *Global Biogeochemical Cycles*, *13*(2), 647–662. https://doi.org/10.1029/1999gb900015
107. Sprent, J. I., & James, E. K. (2007). Legume Evolution: Where Do Nodules and Mycorrhizas Fit In? *Plant Physiology*, *144*(2), 575–581. https://doi.org/10.1104/pp.107.096156
108. Sun, X., Hu, Z., Chen, R., Jiang, Q., Song, G., Zhang, H., & Xi, Y. (2015). Targeted mutagenesis in soybean using the CRISPR-Cas9 system. *Scientific Reports*, *5*(1). https://doi.org/10.1038/srep10342
109. Suzuki, H., Fukushima, E. O., Shimizu, Y., Seki, H., Fujisawa, Y., Ishimoto, M., … Muranaka, T. (2019). Lotus japonicus Triterpenoid Profile and Characterization of the CYP716A51 and LjCYP93E1 Genes Involved in Their

Biosynthesis In Planta. *Plant and Cell Physiology*, *60*(11), 2496–2509. https://doi.org/10.1093/pcp/pcz145
110. Symington, L. S., & Gautier, J. (2011). Double-Strand Break End Resection and Repair Pathway Choice. *Annual Review of Genetics*, *45*(1), 247–271. https://doi.org/10.1146/annurev-genet-110410-132435
111. Tirichine, L., Imaizumi-Anraku, H., Yoshida, S., Murakami, Y., Madsen, L. H., Miwa, H., … Stougaard, J. (2006). Deregulation of a Ca 2+ /calmodulin-dependent kinase leads to spontaneous nodule development. *Nature*, *441*(7097), 1153–1156. https://doi.org/10.1038/nature04862
112. Wang, L., Rubio, M. C., Xin, X., Zhang, B., Fan, Q., Wang, Q., … Duanmu, D. (2019). CRISPR/Cas9 knockout of leghemoglobin genes inLotus japonicusuncovers their synergistic roles in symbiotic nitrogen fixation. *New Phytologist*, *224*(2), 818–832. https://doi.org/10.1111/nph.16077
113. Wang, L., Wang, L., Tan, Q., Fan, Q., Zhu, H., Hong, Z., … Duanmu, D. (2016). Efficient Inactivation of Symbiotic Nitrogen Fixation Related Genes in Lotus japonicus Using CRISPR-Cas9. *Frontiers in Plant Science*, *7*. https://doi.org/10.3389/fpls.2016.01333
114. Weber, E., Engler, C., Gruetzner, R., Werner, S., & Marillonnet, S. (2011). A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLoS ONE*, *6*(2), e16765. https://doi.org/10.1371/journal.pone.0016765
115. Wilson, F. M., Harrison, K., Armitage, A. D., Simkin, A. J., & Harrison, R. J. (2019). CRISPR/Cas9-mediated mutagenesis of phytoene desaturase in diploid and octoploid strawberry. *Plant Methods*, *15*(1). https://doi.org/10.1186/s13007-019-0428-6
116. Xie, K., & Yang, Y. (2013). RNA-Guided Genome Editing in Plants Using a CRISPR–Cas System. *Molecular Plant*, *6*(6), 1975–1983. https://doi.org/10.1093/mp/sst119
117. Xu, R., Fazio, G. C., & Matsuda, S. P. T. (2004). On the origins of triterpenoid skeletal diversity. *Phytochemistry*, *65*(3), 261–291. https://doi.org/10.1016/j.phytochem.2003.11.014
118. Youn, J.-H., & Kim, T.-W. (2015). Functional insights of plant GSK3-like kinases: multi-taskers in diverse cellular signal transduction pathways. *Molecular Plant*, *8*(4), 552–565. https://doi.org/10.1016/j.molp.2014.12.006
119. Yuan, M., Zhu, J., Gong, L., He, L., Lee, C., Han, S., … He, G. (2019). Mutagenesis of FAD2 genes in peanut with CRISPR/Cas9 based gene editing. *BMC Biotechnology*, *19*(1). https://doi.org/10.1186/s12896-019-0516-8
120. Zhang, B., Yang, X., Yang, C., Li, M., & Guo, Y. (2016). Exploiting the CRISPR/Cas9 System for Targeted Genome Mutagenesis in Petunia. *Scientific Reports*, *6*(1). https://doi.org/10.1038/srep20315
121. Zheng, N., Li, T., Dittman, J. D., Su, J., Li, R., Gassmann, W., … Yang, B. (2020). CRISPR/Cas9-Based Gene Editing Using Egg Cell-Specific Promoters in Arabidopsis and Soybean. *Frontiers in Plant Science*, *11*. https://doi.org/10.3389/fpls.2020.00800